Validation of a New 3D/2D Registration Criterion including Error Prediction. Application to Image Guided Radio-Frequency Ablation of the Liver Tumors

Stephane Nicolau, Xavier Pennec, Luc Soler, Nicholas Ayache

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Validation of a New 3D/2D Registration Criterion including Error Prediction. Application to Image Guided Radio-Frequency Ablation of the Liver Tumors

S. Nicolau, X. Pennec, L. Soler, N. Ayache

N° 4993

November 2003

THÈME 3

ISSN 0249-6399 ISRN INRIA/RR--4993--FR+ENG

INRIA
SOPHIA ANTIPOLIS

Validation of a New 3D/2D Registration Criterion including Error Prediction. Application to Image Guided Radio-Frequency Ablation of the Liver Tumors

S. Nicolau, X. Pennec, L. Soler, N. Ayache

Thème 3 — Interaction homme-machine,
images, données, connaissances
Projets Epidaure

Rapport de recherche n° 4993 — November 2003 — 57 pages

Abstract: Our purpose is to provide an augmented reality system for Radio-Frequency (RF) tumor ablation guidance that could superimpose a 3D model of the liver, its vessels and tumors (reconstructed from CT images) on external video images of the patient. The main constraints are the reliability and the accuracy, which justifies a 3D/2D registration based on radio-opaque fiducials rather than surface- or 3D- based techniques. Then, a lack in the statistical assumptions of the classical 3D/2D registration methods led us to the derivation of a new extended 3D/2D criterion. Careful validation experiments on real data show that an accuracy of 2 mm can be achieved in clinically relevant conditions, and that our new criterion is up to 9% more accurate, while keeping a computation time compatible with real-time at 20 to 40 Hz. However, the accuracy of the registration strongly vary w.r.t. the experimental conditions (cameras angle, number of markers...). Thus, to provide a safe system, we should supply an error prediction that take this parameters into account. Propagating the data noise through both our criterion and the classical one, we obtain an explicit formulation of the registration error. As the real conditions do not always fit the theory, it is critical to validate our prediction with real data. Thus, we perform a rigorous incremental validation of each assumption using successively: synthetic data, real video images of a precisely known object, and finally real CT and video images of a soft phantom. Results point out that our error prediction is fully valid in our application range. Eventually, we provide an accurate RA guidance system that allows the automatic detection of potentially inaccurate guidance.

Key-words: Augmented Reality, 3D/2D Registration, Uncertainty Prediction, Tumors Liver Radio-frequencies Ablation

Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles, BP 93, 06902 Sophia Antipolis Cedex (France)
Téléphone : +33 4 92 38 77 77 — Télécopie : +33 4 92 38 77 65

Validation d'un nouveau critère de recalage 3D/2D incluant une prédiction de l'erreur. Application à l'ablation de tumeurs du foie par radio-fréquences assistée par l'image

Résumé : Dans le contexte de l'intervention clinique d'ablation par radio-fréquences, nous souhaitons réaliser un sytème de guidage par réalité augmentée qui superpose les modèles 3D du foie, des vaisseaux et de la tumeur (reconstruits à partir de coupes CT) sur des images vidéo externes du patient. Les contraintes majeures d'un tel système sont la fiabilité et surtout la précision, ce qui justifie notamment l'emploi d'un recalage 3D/2D basé sur des marqueurs radio-opaques plutôt qu'un recalage surfacique, iconique ou 3D/3D. Une lacune dans les hypothèses statistiques des méthodes de recalages classiques 3D/2D nous conduit au développement d'un nouveau critère 3D/2D étendu. Une validation expérimentale sur des données réelles montrent qu'il est possible d'atteindre une précision moyenne de 2 mm dans des conditions pertinentes par rapport à notre application et que notre critère apporte un gain en précision de l'ordre de 9%, tout en conservant des temps de calcul compatibles avec du temps réel à 20-40 Hz. Néanmoins, la précision du recalage peut varier considérablement en fonction des conditions expérimentales (angle entre les caméras, nombre et position relative des marqueurs...). Pour proposer un produit sécurisé, nous devrions fournir une prédiction fiable de l'erreur du système en fonction des différents paramètres possibles. Propageant le bruit des données à travers notre critère et le critère classique, nous obtenons une formulation explicite de l'erreur de recalage en fonction des différents paramètres du système. Comme les conditions réelles ne satisfont pas toujours la théorie, il est primordial de valider notre prédiction sur des données réelles. Nous réalisons pour cela une validation incrémentale rigoureuse de chacune des hypothèses en utilisant successivement : des données synthétiques, des images vidéo d'un objet précisement connu, et finalement des images CT et vidéo réelle d'un mannequin de l'abdomen. Les résultats indiquent que notre prédiction d'erreur est fiable dans l'intervalle des conditions cliniques possibles. Finalement, nous fournissons un système de guidage par réalité augmentée qui permet de détecter automatiquement les configurations du système pouvant mener à un guidage trop imprécis et donc potentiellement dangeureux.

Mots-clés : Réalité augmentée, Recalage 3D/2D, Prédiction d'incertitude, Ablation par radio-fréquence de tumeurs du foie

Contents

1 Introduction

Liver's tumors ablation by radio-frequencies (RF) is a new technique that begins to be widely used by surgeons. However, the guidance procedure to reach the tumors with the electrode is still made visually using per-operative cross-sections of the patient. The visualization can be made either by Ultra-Sound (US) or Computed Tomography (CT) images, but in both cases the surgeon must cope with the problem of interpreting 2D-slices to locate the tumor's center in the 3D space. This is the reason why surgeons estimate they cannot make a reliable intervention if the tumor's diameter does not exceed 2 cm.

Our purpose is to build an augmented reality system that could superimpose the 3D liver and tumor reconstructions of the patient on video images. This would provide to surgeons information about the position of the tumor in order to improve their accuracy during the guidance step. Hence the accuracy of the guiding system is one of our main concern. Indeed the overall accuracy of the guiding system has to be less than 5 mm to be a significant help for the surgeon performing his task.

Just before the intervention, a CT-scan of the patient is acquired and the 3D-reconstructions of his skin, his liver and the tumors is performed [Sa01]. Two cameras (jointly calibrated) are viewing the patient's skin from two different points of view. The patient is intubated during the intervention, so the volume of gas in his lungs can be controlled and monitored. Then, it is possible to fix the volume at the same value during a few seconds repetitively and to perform the CT and the electrode's manipulation almost in the same volume's condition. Balter [BLM+98] indicates that the mean tumor repositioning at exhalation phase in a respiratory-gated radiotherapy context is under 1 mm. Thus, it is reasonable to assume that a rigid registration is sufficient to register accurately the 3D-model extracted from the CT with the 2D video images. Consequently we are confronted to the classical rigid problem of *3D/3D and 3D/2D Object Registration.*

1.1 3D/3D and 3D/2D Registration Methods

There are three main classes of methods to solve this registration problem, quickly described below.

Iconic registration needs only one video image and a 3D-reconstruction of the object. The registration method is based on the assumption that it exists a statistical relationship between the grey level in the video image and the normal vectors of the

3D model's surface. Then, this relationship is introduced in a criterion which involves mutual information calculations.

Surface registration needs an apparatus that can acquire 3D-coordinates of a large amount of points on the object to register (in our case the skin of the patient). This apparatus can be a stereoscopic reconstruction provided by two cameras or a laser range scanner. In certain cases, an extra calibration step can be necessary to determine accurately the relationship between the cameras reference frame and the apparatus reference. Then, we are left with two surface to register that contains lots of points (generally 1000 to 100000). Usually the registration is computed with an ICP algorithm.

Landmarks registration is based on the possibility to find several points that are both visible on the 3D-model and in the video images. Then, it is possible to make either a 3D/3D or a 2D/3D registration. In the first case, the 3D coordinates of the visible landmarks need to be computed (for instance by stereoscopy), and the registration between the two 3D points sets is the result of the classical 3D/3D LSQ criterion optimization. In the second case, the usual method consists in minimizing the sums of the squares of the distances between the projected landmarks of the 3D model with their detected position in the video images.

1.2 Discussion

Iconic registration has been proposed to register the surface of 3D objects to video images [Vio95, VW97], with application to the skin surface of the face. The computation time was equivalent to 1 sec. on a current 500 MHz PC. The accuracy reported in [Lev97, LWG97] was 3 mm (RMS) for an object of 50 cm viewed by two camera at a distance of 1 meter with an angle between the cameras of about 70 degrees. However, the 3D object used was a car model, thus presenting more curved surfaces than a human face, and the methods requires particular illumination conditions on the skin that would probably not be fulfilled in our case [Her98]. Moreover, the cylindrical shape of the human chest is likely to lead to very large uncertainties along the cranio-caudal axis. Thus, it seems difficult to obtain good enough results from this method in our case.

Surface registration has been used for augmented reality [GEW+96] in the context of neuro-surgical operations. The surface of the skin extracted from an MR image was registered to a set of points acquired on the face with a laser range scanner. Thanks to several highly curved "edges" on the face (nose, ears, eyes), the reported

accuracy was around 1 mm. One could think that registering many points on surfaces could lead to a better accuracy, proportional to $\sigma/\sqrt{n}$ (where σ is the variance on the point and n the number of points), but this is usually not the case since point measurements are correlated and are not real landmarks (two samplings of a surface measure different physical points). In our case, the particular geometry of the chest will moreover disturb the quality of the registration. As the chest does not have very curved parts, the translation along the body axis is very loosely constrained, despite an important density of point to be registered.

Thus, we think that a registration based on landmarks is better adapted to our problem. Indeed, computation times are usually very low (under 0.01 sec.) and the accuracy seems to be sufficient, even if there are very few clinical evaluations (0.5 to 1.4 mm in the case of 3D/3D registration of head implanted markers [MFW+97]). In our case, there are no visible anatomical landmarks, so we chose to stick some radio-opaque markers to the patient skin. Using a stereoscopic reconstruction to perform a 3D/3D registration would lead us to select only the markers visible in the two camera images. This would reduce the achievable accuracy (since there are less data used to compute the registration) and could moreover add some systematic or random errors due to the reconstruction algorithm. Thus, we opted for a 2D/3D registration. The current procedure is to localize interactively the fiducials centers on the 3D CT image and on the video images (automatic segmentation algorithms are being evaluated). The matching is performed thanks to epipolar geometry between video points, and using a prediction/verification (alignment) algorithm between video and CT points. This time, as our points are really homologous, the "cylindrical" shape of the chest is not any more a problem.

1.3 Overview of the Document

In the sequel, we focus on the 2D/3D registration problem and its performance evaluation. More particularly, we concentrate on the criterion to optimize for the rigid transformation, and on the evaluation of its optimum performances without much attention to the optimization technique.

We detail in Section 2 the classical 2D/3D registration criterion for points correspondences (referred from now on to the *Standard Projective Points Correspondences* for SPPC), and we show that it can be considered as a *Maximum Likelihood* estimation if the 3D points are exact (Sec. 2.3.1). Then, modifying the statistical assumptions to account for an existing noise on the 3D point measurements, we derive in Section 2.3.2 an *Extended* and original criterion (EPPC) that generalizes the

classical one. Section 3 is devoted to a perturbation analysis of the two previous criteria to predict the uncertainty of the resulting transformation.

In Sec. 4, we estimate and compare the performances of both criteria in terms of speed, robustness and accuracy. This study on synthetic and real data proves that EPPC provides a registration up to 20% more accurate than SPPC and that an accuracy of 2 mm can be reached in clinical conditions. In Sec. 5 we outline the necessity to validate our error prediction since our perturbation analysis is done under particular assumptions. Therefore, we carry out a rigorous validation of our error prediction and show incrementally that each hypothesis is verified.

2 A new 2D/3D Registration Criteria

To provide fiducials for the 3D/2D registration, we stick radio-opaque markers on the patient's skin in such a way that they can be seen by the cameras. We are then confronted to the classical problem of an *Object Pose from 2D to 3D Points Correspondences*.

2.1 Notations

Let $M_i(x_i, y_i, z_i)$ be the 3D points that represent the exact localization of the radio-opaque fiducials in the CT-scan reference frame and $m_i^{(l)}(u_i, v_i)$ be the 2D points that represent its exact position (in floating units) in the video images. The index (l) indicates whose camera the point $m_i^{(l)}$ is belonging to. In this section, we assume that correspondences have already been done, i.e. M_i is matched to $m_i^{(l)}$. To account for occlusion, we use a binary variable ξ_i^l equal to 0 if M_i is out of the field of view or occluded for camera (l) and 1 if it is observed.

We denote by $T \star M$ the action of the rigid transformation T on the 3D point M. Let P_l $(1 \leq l \leq M)$ be the camera's projective functions from 3D to 2D. That means we have these projective relationships:

$$P^{(l)}(T \star M_i) = m_i^{(l)} \tag{1}$$

In the following sections, $\hat{A}$ will represent an *estimation* of a perfect data A, and $\tilde{A}$ an *observed measure*. Thus, the 3D points measured by the user will be written $\tilde{M}_i$, and the measured video 2D points $\tilde{m}_i$.

2.2 Related Works

This problem was largely considered in a wide variety of case. Pioneer works considered less than 6 matched points and derived a closed-form solution valid in most geometrical cases [FB81, DRL89, HYH85, QL99]. These methods are direct but they are very sensitive to noise because the transformation is derivated from exact geometric relationships between the minimal number of points. When the number of available points correspondences is larger than the number of unknowns, it is preferable to use all the points in order to minimize the influence of their measurement error, and to take all the information into account.

Ganapathy and others [Gan84, LHF90] proposed a linear resolution by replacing the rotation by a 3×3 matrix without any constraints. This means they approximate the rigid transformation by an affine one. This linear method presents the advantage of solving directly the pose calculation. However, it is very sensitive to noise mainly because of the absence of constraints undergone by the rotation matrix. DeMenthon [DD95] proposed a gracefully linear method approximating the *perspective projection* with a *scaled orthographic projection*. The direct result, supplied by his linear formulation of the problem, is then followed by an iterative process improving the initial solution. More recently, Quan [QL99] presented a straight linear solution for 4 points and more, based on inter-points distance constraints. His problem's formulation lead to a fourth degree polynomial system that he solves linearly by decreasing constraints on the unknowns.

Yuan [Yua89] and Liu [LHF90] solve the problem by separating the rotational components from the translational one. Yuan demonstrates that, given 3D to 2D points correspondences, the rotation parameters are the common roots of six quadratic equations. Then, they are found using Newton's iterative gradient descent. In the case of 3D to 2D lines correspondences, Liu finds the rotational components iteratively by modeling the rotation with Euler's angle.

Lowe, Phong and Simon [Low91, PHT95, SB97] determine the object pose by minimizing a classical least-squares (LSQ) error function representing the sums of the squares of the distances between the projected points with their detected position in the video image. In these cases, the methods differ by the minimization method (Newton descent, Powell's algorithm, trust-region method), and by the rotation parameterization (rotation vector, euler angles, dual quaternions).

Haralick [HJL$^+$89] and Or [OLWK98] turn the 2D/3D problem into a 3D/3D points registration problem in which they estimate the depth of the points seen in the image. They both minimize a 3D LSQ Euclidean distance criterion in two steps. Or uses a *Maximum Likelihood* formulation to describe his criterion. This

means that he iteratively estimates the transformation and the depths of the points. On the other hand, Haralick provides an iterative calculation of the depths based on geometric consideration of the system. Moreover, each term of the criterion is weighted by its observation's goodness.

Let us now discuss how this huge number of different proposition to solve the pose estimation problem do adapt to our application requirements. Most of the linear solution were motivated by real-time application and, up to the nineties, the computations times needed to optimize a LSQ criterion were prohibitive. Currently, it can be performed in less than 0.1 sec. for 30 points with an average initialization. Therefore, as the accuracy is crucial in our application, we cannot afford using a linear method that is more sensitive to noise.

The alternative approach of separating the rotation from the translation was examined by Kumar [Kum89], and he shows that it leads to worse parameters estimation in the presence of noise than the classical LSQ estimation. Unfortunately, no comparison reference have been published about DeMenthon's method, therefore it is not possible to criticize its accuracy performance.

Nevertheless, we think that a LSQ criterion has a definite advantage among the other methods because it is possible to predict accurately the influence of the noise on the final registration and therefore, on the points to be registered [PT97]. For our application, we believe that it is even crucial since it would allow the detection of any bad configuration of markers that would lead to an untruthful registration.

Such an uncertainty prediction method currently exists for 3D/3D registration, but, up to our knowledge, this is the first time that an analytical formulation is presented for the 3D/2D case (Section 3). Last but not least, all of the existing methods implicitly consider that 2D points are noisy, but that 3D points are exact. In our case, this assumption is definitely questionable, which lead to the development of a new maximum likelihood (ML) criterion generalizing the standard 3D/2D LSQ criterion in Section 2.3.2. Interestingly, this new criterion can also be viewed as a generalization of the 3D reconstruction followed by a 3D/3D registration technique (Sec. 2.4).

2.3 Maximum Likelihood 2D/3D Registration

2.3.1 Standard Projective Points Correspondences (SPPC) criterion

In this section, we reformulate the classical 3D/2D LSQ criterion as a *Maximum Likelihood (ML)* estimation. This exercise allows us to derive the criterion as a

consequence of the statistical noise assumptions on our data, and will be the basis for the derivation of the more general ML algorithm of Section 2.3.2.

Let us assume that our 3D points are exact (i.e. the measured 3D points $\tilde{M}_i$ are equal to the exact 3D points M_i) and that only the 2D points are corrupted by an isotropic Gaussian noise η_i of covariance matrix

$$\Sigma_{2D} = \begin{pmatrix} \sigma_{2D}^2 & 0 \\ 0 & \sigma_{2D}^2 \end{pmatrix}$$

Taking into account the geometric relations between exact 3D and 2D points (Eq. 1), the noise model for the measured 2D points becomes:

$$\tilde{m_i}^{(l)} = m_i^{(l)} + \eta_i = P^{(l)}(T \star M_i) + \eta_i \quad \text{with} \quad \eta_i \sim N(0, \sigma_{2D})$$

Thus, the probability density of measuring the projection of the 3D point $M_i = [x_i, y_i, z_i]^\top$ at the location $\tilde{m_i}^{(l)} = [u_i^{(l)}, v_i^{(l)}]^\top$ in image (l), knowing the (transformation) parameters $\theta = \{T\}$ is given by:

$$p(\tilde{m_i}^{(l)} \mid \theta) = k_{2D} \cdot \exp\left(- \frac{\| P^{(l)}(T \star M_i) - \tilde{m_i}^{(l)} \|^2}{2{\sigma_{2D}}^2} \right) = G_{\sigma_{2D}}(P^{(l)}(T \star M_i) - \tilde{m_i}^{(l)}),$$

where $k_{2D} = 1/2\pi\sigma_{2D}^2$ is the normalization constant.

Let χ be our data vector (regrouping all the measured data, in this case the 2D points $\tilde{m_i}^{(l)}$ only, since the 3D points are assumed to be exact). Since the detection of each 2D point is performed independently, the probability of the observed data is simply $p(\chi \mid \theta) = \prod_{l=1}^{M} \prod_{i=1}^{N} p(\tilde{m_i}^{(l)} \mid \theta)^{\xi_i^l}$. In this formula, unobserved 2D points (for which $\xi_i^l = 0$) are implicitly taken out of the probability.

Now, the *Maximum likelihood* transformation is the transformation $\hat{\theta}$ that maximizes the probability of the observed data, or equivalently, that minimizes its negative log:

$$\hat{\theta} = \arg\min_{\theta} \; -log[p(\chi \mid \theta)]$$

Developing the likelihood function gives:

$$-log[p(\chi \mid \theta)] = -log\Big[\prod_{l=1}^{M} \prod_{i=1}^{N} p(\tilde{m_i}^{(l)} \mid \theta)^{\xi_i^l} \Big] = \sum_{l=1}^{M} \sum_{i=1}^{N} -log\Big[p(\tilde{m_i}^{(l)} \mid \theta)^{\xi_i^l} \Big]$$

Hence, we end up by minimizing

$$C_{2D}(T) = \sum_{l=1}^{M} \sum_{i=1}^{N} \xi_i^l \cdot \frac{\left\| P^{(l)}(T \star M_i) - \tilde{m}_i^{(l)} \right\|^2}{2 \cdot {\sigma_{2D}}^2} - \left(\sum_{l=1}^{M} \sum_{i=1}^{N} \xi_i^l \right) \cdot log[k_{2D}] \qquad (2)$$

Thus, up to a constant and a multiplicative factor, this ML estimation of the transformation boils down to the classical LSQ criterion in the 2D coordinates. Of course, this criterion assumes that there is no noise on the 3D points, the idea being that the possible 3D noise could be distributed over the 2D measurements. This simple hypothesis providing a smart quadratic criterion with only 6 parameters to estimate, leads obviously to a very low calculations cost (less than 0.01 s). However, from a statistical point of view, distributing the 3D error on 2D measurements leads to correlated noises, which does not agree with the independence assumption used to derive le ML estimation.

2.3.2 Extended Projective Points Correspondences (EPPC) criterion

To introduce a possible noise on the 3D data, it is thus safer to consider that we are measuring a noisy version of the exact points (once again, we consider here an isotropic and Gaussian noise):

$$\tilde{M}_i = M_i + \varepsilon_i \qquad \text{with} \qquad \varepsilon_i \sim N(0, \sigma_{3D})$$

In this case, the exact location M_i of the 3D points is considered as a parameter, just as the transformation T. In statistics, this is called a *latent or hidden variable*, while it is better known as an *auxiliary variable* in computer vision. Thus, knowing the parameters $\theta = \{T, M_1, \dots M_N\}$, the probability of measuring a 3D point is:

$$p(\tilde{M}_i \mid \theta) = G_{\sigma_{3D}}\left(M_i - \tilde{M}_i\right), \tag{3}$$

and the probability of measuring a 2D point is still:

$$p(\tilde{m}_i^{(l)} \mid \theta) = G_{\sigma_{2D}}\left(P^{(l)}(T \star M_i) - \tilde{m}_i^{(l)}\right) \tag{4}$$

One important feature of this statistical modeling is that we can safely assume that all 3D and 2D measurements are independent[1]. Thus, we can write the probability of our observation vector $\chi = (\tilde{m}_1^1, ..., \tilde{m}_N^1, ..., \tilde{m}_1^M, ..., \tilde{m}_N^M, \tilde{M}_1, ..., \tilde{M}_N)$ as the product of individual probabilities:

$$p(\chi \mid \theta) = \left(\prod_{l=1}^{M}\prod_{i=1}^{N} p(\tilde{m_i}^{(l)} \mid \theta)^{\xi_i^l}\right) \cdot \left(\prod_{i=1}^{N} p(\tilde{M}_i \mid \theta)\right) \tag{5}$$

[1] More precisely, the 3D and 2D point measurements are conditionally independent given the parameters $\theta = \{T, M_1, \dots M_N\}$.

The ML estimation of the parameters is still optimizing the probability of the observed data. Taking the negative log, we obtain $C(\theta) = C(T, M_1, \dots M_N) = -log(p(\chi \mid \theta))$. Plugging the values from the equations 3, 4, 5 above, we get:

$$C(T, M_1, \dots M_N) = \sum_{i=1}^{N} \frac{\| \tilde{M}_i - M_i \|^2}{2 \cdot {\sigma_{3D}}^2} + \sum_{l=1}^{M} \sum_{i=1}^{N} \xi_i^l \cdot \frac{\| \tilde{m}_i^{(l)} - m_i^{(l)} \|^2}{2 \cdot {\sigma_{2D}}^2} + K(\sigma_{2D}, \sigma_{3D}) \tag{6}$$

where $m_i^{(l)} = P^{(l)}(T \star M_i)$ and K is a normalization constant depending on σ_{2D} and σ_{3D}. The convergence is insured since we minimize the same positive criterion at each step.

The obvious difference between this criterion and the SPPC is that we have additional unknowns in the system to be solved (the hidden variables). It is then necessary to modify the optimization algorithm. Indeed, there are now two ensembles of unknowns to estimate, the first one represents the rigid transformation and the second one corresponds to the M_i.

Among the multiple possibilities, an obvious choice is to perform an alternated minimization of the criterion along the two groups of variables. At the first step, we initialize the M_i with the $\tilde{M}_i$ and perform a first minimization on T (using a first initialization T_0). In fact, this corresponds to optimizing the SPPC criterion. At the second step, we keep the transformation T fixed and we optimize for the M_i. We then continue to alternatively update the transformation $\hat{T}$ (given the lastly estimated values $\hat{M}_i$ of the exact M_i) and the exact positions $\hat{M}_i$ (given the lastly estimated transformation $\hat{T}$). The algorithm is stopped when the distance between the two last estimations of T becomes negligible:

$$dist(\hat{T}^{i+1}, \hat{T}^i) < d_T$$

where $dist(\cdot, \cdot)$ is a 2×1 vector $[d_r, d_t]^\top$ containing the rotation and translation vector norms of the composition $\hat{T}^{i+1} \cdot (\hat{T}^i)^{-1}$, d_T is a 2×1 vector $[d_{max}^r, d_{max}^t]^\top$ chosen with respect to the seeked accuracy on the transformation[2]. To be rigorous, we should take into account the second part of the data to decide when stopping the alternated minimization (i.e. the evolution of the M_i), but as we are interested only by the transformation, we can neglect it.

We can notice that it would not be foresighting to use only the value of C to decide when stopping the alternated iteration. Indeed, we do not know the true values of σ_{3D}^2 and σ_{2D}^2, which means that we only have a rough estimation of the

[2]In this context $<$ means: $d_r < d_{max}^r \quad AND \quad d_t < d_{max}^t$

real value of C and that we are not able to know when we are sufficiently close to the algorithm convergence.

2.4 Link with Reconstruction and 3D/3D Registration

Let us consider the change of variable $M_i' = T \star M_i$. This amount to consider that the exact 3D points are measured in the reference frame of the cameras (instead of the CT frame as previously) and that we are looking for the transformation $T' = T^{(-1)}$ (the transformation from the camera world to the CT instead of the the transformation from the CT to the cameras). Dropping the primes, The 3D/2D ML criterion is then rewritten

$$C(T, M_1, \dots M_N) = \sum_{i=1}^{N} \frac{\| \tilde{M}_i - T * M_i \|^2}{2 \cdot {\sigma_{3D}}^2} + \sum_{l=1}^{M} \sum_{i=1}^{N} \xi_i^l \cdot \frac{\| \tilde{m}_i^{(l)} - P^{(l)}(M_i) \|^2}{2 \cdot {\sigma_{2D}}^2} + K$$

As explained in the section 2.3.2, this criterion can be optimized iteratively by successively estimating the 3D coordinates and the transformation T. Moreover, it has to be noticed that the change of variable does not affect the transformation to be found.

Now, assuming that we are performing an estimation with a largely overestimated σ_{3D} and a correct estimation of σ_{2D}. Around the optimal transformation $\hat{T}$, we will have

$$\frac{1}{N} \sum_{i=1}^{N} \| \tilde{M}_i - T * M_i \|^2 \simeq \frac{1}{N} \sum_{i=1}^{N} \| \tilde{M}_i - \hat{T} * M_i \|^2 = \hat{\sigma}_{3D} \ll \sigma_{3D}$$

Thus, the first term of the criterion is negligible with respect to the second term (since σ_{2D} is assumed to be correctly estimated): optimizing for the exact 3D point positions boils down to the minimization of

$$C_{Rec}(M_1, \dots M_N) = \sum_{l=1}^{M} \sum_{i=1}^{N} \xi_i^l \cdot \| \tilde{m}_i^{(l)} - P^{(l)}(M_i) \|^2$$

This criterion is in fact one of the more widely used reconstruction criterion. Then, after the determination of the 3D coordinates of the points in the cameras frame, the next step consists in minimizing the criterion with respect to T. As the second term does not depends on T, this corresponds to the minimization of

$$C_{3Dreg}(T) = \sum_{i=1}^{N} \| \tilde{M}_i - T * M_i \|^2$$

which is no more than the standard 3D LSQ criterion.

As a conclusion, the method consisting in reconstructing the position of 3D points from the cameras and then registering in 3D can be viewed as a limit case of our 3D/2D ML criterion where the noise on 3D points is largely *overestimated* (with respect to the noise on 2D points). On the other hand, we have already seen that the standard 2D ML criterion for 3D/2D registration is a limit case of our 3D/2D ML criterion where the noise on 3D points is largely *underestimated* or really very small (still with respect to the noise on 2D points).

Thus, we may expect our criterion to perform better than these methods when there is effectively some noise on the 3D points and if we have a good estimation of relative 3D and 2D variances.

3 Uncertainty Prediction

In the surgical setup, during the AR-guided radio-frequency intervention, the 3D/2D registration will be used to locate the target within the patient and to guide the introduction of the electrode. From a safety point of view, one of the main concern is to ensure that the registration is accurate enough to guaranty that only the planned target is reached by the electrode. We will see in Sec. 4 that the accuracy provided can sharply vary w. r. t. the initial conditions (the number of markers, the angle between the cameras, the focus, and the relative position of the target w.r.t. the markers). For instance there can be a factor two on the accuracy when the cameras angle goes from 20^o to 60^o. In accordance with this fact, we cannot afford providing a system without assessing its accuracy during any possible intervention.

In order to propose a safe product to surgeons, we should provide a statistical study that would give the mean *Target Registration Error* (TRE) with respect to the possible input parameters. This is the equivalent of the direction for use and the secondary effects list mandatory for all proposed drugs in the therapeutic field, and the reliability and accuracy tables of robotics tools: these tables give a usability range to assess under which condition a particular feature (for example accuracy) could be reached.

As increasing the number of experiments is very expensive and time-consuming, it is almost infeasible to *measure* the accuracy provided for each experimental condition. Moreover, as we want a real-time system, the conditions may change during the operation (e.g. markers can be occluded by the surgeon), and the accuracy assessment has to be constantly updated to avoid a potentially dangerous gesture.

Consequently, we think that predicting the TRE by studying the theoretical noise propagation is the best way to ensure the safety of our system.

3.1 Basic Knowledge

We briefly remind here the calculations steps that lead to the transformation's covariance from the data covariance and a criterion involving the data and the transformation (for more details see [Pen96]).

Let us assume that the criterion $C(\chi, T)$ is a function of class C^2 from $\mathbb{R}^m \times \mathbb{R}^p$ to $\mathbb{R}^+$, where χ is the m-dimensional vector of data, and T the p-dimensional vector of sought parameters (in our case the transformation). The optimal transformation $\hat{T}$ is defined as the one that minimizes the criterion: $\hat{T} = \arg\min_T(C(\chi, T))$. A local minima is reached and well defined if and only if:

$$\Phi(\chi, T) = \left(\frac{\partial C}{\partial T}(\chi, T)\right)^\top = 0 \qquad \text{and} \qquad H = \frac{\partial^2 C}{\partial T^2}(\chi, T) \qquad \text{positive definite.} \quad (7)$$

The function Φ implicitly define $\hat{T}$ as a function of the data χ. A first order Taylor expansion of Φ gives us:

$$\Phi(\chi + \delta\chi, T + \delta T) = \Phi(\chi, T) + \frac{\partial \Phi}{\partial \chi} \cdot \delta\chi + \frac{\partial \Phi}{\partial T} \cdot \delta T + O(\delta\chi^2, \delta T^2)$$

which means that around an optimum $\hat{T}$ we have:

$$\hat{T}(\chi + \delta\chi) = \hat{T}(\chi) - \left(\frac{\partial \Phi}{\partial T}\right)^{(-1)} \cdot \frac{\partial \Phi}{\partial \chi} \cdot \delta\chi + O(\delta\chi^2)$$

This result may also be obtained using the implicit function theorem. Thus, if χ is a random vector of mean $\bar{\chi}$ and covariance $\Sigma_{\chi\chi}$, the optimal transformation $\hat{T}$ is (up to the second order), a random vector with the following mean and covariance:

$$\bar{T} = \arg\min_T(C(\bar{\chi}, T)) \qquad \text{and} \qquad \Sigma_{TT} = H^{-1} \cdot \left(\frac{\partial \Phi}{\partial \chi}\right) \cdot \Sigma_{\chi\chi} \cdot \left(\frac{\partial \Phi}{\partial \chi}\right)^\top \cdot H^{-T}$$

Thus, to propagate the covariance matrix from the data to the parameters of the transformation optimizing the criterion, we need to compute $H = \frac{\partial^2 C(\chi,T)}{\partial T^2}$ and $J_{\Phi_x} = \frac{\partial^2 C(\chi,T)}{\partial \chi \partial T}$. To obtain the final covariance matrix on a target point C_i after registration, we simply have to propagate the uncertainty through the transformation action:

$$\Sigma_{T \star C_i} = \frac{\partial (T \star C_i)}{\partial T} \cdot \Sigma_{TT} \cdot \frac{\partial (T \star C_i)}{\partial T}^\top \quad (8)$$

3.2 Theoretical Uncertainty of the SPPC

The SPPC, developed in Sec. 2.3.1, corresponds to the ML estimation of the transformation with exact 3D points M_i (we consider in this section that $M_i = \tilde{M}_i$) and noisy 2D points $\tilde{m_i}^{(l)}$. However, to account for all possible sources of uncertainty, we need to investigate how does the solution of this criterion vary if there is some noise on the 2D data, but also on the 3D data. Thus, we need to include the M_i's is the data and to assume a covariance matrix Σ_{M_i} on each 3D measurement. The criterion to minimize is given by Eq. 2:

$$C(T) = \sum_{l=1}^{M} \sum_{i=1}^{N} \xi_i^l \cdot \frac{\| P^{(l)}(T \star M_i) - \tilde{m_i}^{(l)} \|^2}{2\sigma_{2D}^2}$$

To reduce the complexity of the calculations, we decompose the projection matrix 3×4 as shown below:

$$P^{(l)} = \left[\begin{array}{c|c} Q_{2\times 3}^{(l)} & b_{2\times 1}^{(l)} \\ \hline C_{(l)}^{\top} & 1 \end{array}\right]$$

We can now write analytically the incidence of the projection function $P^{(l)}$ on a 3D point:

$$m_i^{(l)} = P^{(l)}(T \star M_i) = \frac{Q^{(l)} \cdot (T \star M_i) + b^{(l)}}{1 + C_{(l)}^{\top} \cdot (T \star M_i)},$$

that we can now differentiate w.r.t. the parameters of the transformation:

$$\frac{\partial m_i^{(l)}}{\partial T} = \left(\frac{Q^{(l)}}{1+<C_{(l)}|T\star M_i>} - \frac{\left(Q^{(l)} - m_i^{(l)} \cdot C_{(l)}^{\top}\right) \cdot C_{(l)}^{\top}}{(1+<C_{(l)}|T\star M_i>)^2} \right) \cdot \frac{\partial (T\star M_i)}{\partial T} = \frac{Q^{(l)} - m_i^{(l)} \cdot C_{(l)}^{\top}}{1+ < C_{(l)}|T \star M_i >} \cdot D_i \tag{9}$$

where $D_i = \frac{\partial (T\star M_i)}{\partial T}$ denotes the derivative of the action of the transformation T w.r.t. the parameters of the transformation (see [Pen96] for its expression using the rotation and translation vectors). Likewise, we will need the derivative of $m_i^{(l)}$ with respect to the 3D point M_i:

$$\frac{\partial m_i^{(l)}}{\partial M_i} = \left(\frac{Q^{(l)}}{1+<C_{(l)}|T\star M_i>} - \frac{\left(Q^{(l)} - m_i^{(l)} \cdot C_{(l)}^{\top}\right) \cdot C_{(l)}^{\top}}{(1+<C_{(l)}|T\star M_i>)^2} \right) \cdot \frac{\partial (T\star M_i)}{\partial M_i} = \frac{Q^{(l)} - m_i^{(l)} \cdot C_{(l)}^{\top}}{1+ < C_{(l)}|T \star M_i >} \cdot R \tag{10}$$

where R is the 3×3 rotation matrix of the transformation T such that $T \star M = R \cdot M + t$. To clarify some heavy expressions we define now the following matrix that we will often recognize in our formula:

$$\begin{aligned}
G_i^{(l)} &= (Q^{(l)} - m_i^{(l)} \cdot C_{(l)}^\top)^\top \cdot (Q^{(l)} - m_i^{(l)} \cdot C_{(l)}^\top) - C_{(l)} \cdot (m_i^{(l)} - \tilde{m}_i{}^{(l)})^\top \cdot (Q^{(l)} - m_i^{(l)} \cdot C_{(l)}^\top) \\
&\quad -(Q^{(l)} - m_i^{(l)} \cdot C_{(l)}^\top)^\top \cdot (m_i^{(l)} - \tilde{m}_i{}^{(l)}) \cdot C_{(l)}^\top \\
L_i^{(l)} &= \frac{(Q - m_i^{(l)} \cdot C_{(l)}^\top)^\top \cdot (Q - m_i^{(l)} \cdot C_{(l)}^\top)}{\sigma_{2D}^2 \cdot (1 + < C_{(l)} | T \star M_i >)^2} \qquad \text{and} \qquad L_i = \textstyle\sum_{l=1}^M \xi_i^l \cdot L_i^{(l)} \\
K_i^{(l)} &= \frac{G_i^{(l)}}{\sigma_{2D}^2 \cdot (1 + < C_{(l)} | T \star M_i >)^2} \qquad \text{and} \qquad K_i = \textstyle\sum_{l=1}^M \xi_i^l \cdot K_i^{(l)}
\end{aligned} \tag{11}$$

According to the paragraph 3.1, we have firstly to calculate the implicit function $\Phi = \frac{\partial C}{\partial T}^\top$:

$$\frac{\partial C}{\partial T} = \frac{\partial}{\partial T} \left(\sum_{l=1}^{M} \sum_{i=1}^{N} \xi_i^l \cdot \frac{\| P^{(l)}(T \star M_i) - \tilde{m}_i{}^{(l)} \|^2}{2\sigma_{2D}^2} \right) = \frac{1}{\sigma_{2D}^2} \sum_{l=1}^{M} \sum_{i=1}^{N} \xi_i^l \cdot (m_i^{(l)} - \tilde{m}_i{}^{(l)})^\top \cdot \frac{\partial m_i^{(l)}}{\partial T}$$

Now, using the derivative of $m_i^{(l)}$ given is Eq. 9, we obtain:

$$\Phi = \left(\frac{\partial C}{\partial T} \right)^\top = \frac{1}{\sigma_{2D}^2} \sum_{l=1}^{M} \sum_{i=1}^{N} \xi_i^l \cdot D_i{}^\top \frac{\left(Q^{(l)} - m_i^{(l)} . C_{(l)}^\top \right)^\top \cdot \left(m_i^{(l)} - \tilde{m}_i{}^{(l)} \right)}{1 + < C_{(l)} | T \star M_i >}$$

Then, differentiating once again w.r.t. T, we obtain (after simplification) the Hessian matrix:

$$H = \frac{\partial^2 C}{\partial T^2} = \frac{\partial \Phi}{\partial T} = \frac{1}{\sigma_{2D}^2} \sum_{l=1}^{M} \sum_{i=1}^{N} \xi_i^l \cdot \frac{D_i{}^\top . G_i^{(l)} . D_i}{(1 + < C_{(l)} | T \star \tilde{M}_i >)^2} = \sum_{i=1}^{N} D_i{}^\top \cdot K_i \cdot D_i \tag{12}$$

To compute the Jacobian J_Φ, we need to differentiate Φ with respect both the M_i's and the $\tilde{m}_i^{(l)}$'s:

$$\frac{\partial \Phi}{\partial \tilde{m}_i{}^{(l)}} = -\xi_i^l * \frac{1}{\sigma_{2D}^2} \frac{D_i{}^\top . \left(Q^{(l)} - m_i^{(l)} . C_{(l)}^\top \right)^\top}{1 + < C_{(l)} | T \star M_i >} \tag{13}$$

$$\frac{\partial \Phi}{\partial M_i} = \sum_{l=1}^{M} \xi_i^l \cdot \frac{1}{\sigma_{2D}^2} \frac{D_i{}^\top \cdot G_i^{(l)} \cdot R}{(1 + < C_{(l)} | T \star M_i >)^2} = D_i{}^\top \cdot K_i \cdot R \tag{14}$$

We can now form the Jacobian of Φ:

$$J_\Phi = \frac{\partial \Phi}{\partial \chi} = \left[\frac{\partial \Phi}{\partial M_1}, ..., \frac{\partial \Phi}{\partial M_N}, \frac{\partial \Phi}{\partial \tilde{m}_1^1}, ..., \frac{\partial \Phi}{\partial \tilde{m}_N^1}, ..., ..., \frac{\partial \Phi}{\partial \tilde{m}_1^M}, ..., \frac{\partial \Phi}{\partial \tilde{m}_N^M} \right]$$

where χ is a big vector containing all our data:

$$\chi = \begin{vmatrix} M_1 \\ \vdots \\ M_N \\ \tilde{m}_1^1 \\ \vdots \\ \tilde{m}_N^1 \\ \vdots \\ \tilde{m}_1^M \\ \vdots \\ \tilde{m}_N^M \end{vmatrix} \quad \text{with} \quad \Sigma_{\chi\chi} = \begin{bmatrix} \Sigma_{M_1} & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & & & & & & & \vdots \\ \vdots & \ddots & \Sigma_{M_N} & \ddots & & & & & & \vdots \\ \vdots & & \ddots & \Sigma_{\tilde{m}_1^1} & \ddots & & & & & \vdots \\ \vdots & & & \ddots & \ddots & \ddots & & & & \vdots \\ \vdots & & & & \ddots & \Sigma_{\tilde{m}_1^N} & \ddots & & & \vdots \\ \vdots & & & & & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & & & & & \ddots & \Sigma_{\tilde{m}_1^M} & \ddots & \vdots \\ \vdots & & & & & & & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & \Sigma_{\tilde{m}_N^M} \end{bmatrix}$$

According to Sec. 3.1, the covariance matrix of the transformation parameters is now $\Sigma_{TT} = H^{-1} \cdot \frac{\partial \Phi}{\partial \chi} \cdot \Sigma_{\chi\chi} \cdot \frac{\partial \Phi}{\partial \chi}^\top \cdot H^{-T}$. To simplify the calculations, we reduce firstly $\Gamma = \frac{\partial \Phi}{\partial \chi} \cdot \Sigma_{\chi\chi} \cdot \frac{\partial \Phi}{\partial \chi}^\top$. Assuming an isotropic and identical noise on all 3D points and on all 2D points (i.e. $\Sigma_{M_i} = \sigma_{3D}^2 \cdot \mathrm{Id}_3$ and $\Sigma_{\tilde{m}_i^{(l)}} = \sigma_{2D}^2 \cdot \mathrm{Id}_2$), we have:

$$\begin{aligned} \Gamma &= \sum_{i=1}^{N} \frac{\partial \Phi}{\partial M_i} . \Sigma_{M_i} . \frac{\partial \Phi}{\partial M_i}^\top + \sum_{l=1}^{M} \sum_{i=1}^{N} \frac{\partial \Phi}{\partial \tilde{m}_i^{(l)}} . \Sigma_{\tilde{m}_i^{(l)}} . \frac{\partial \Phi}{\partial \tilde{m}_i^{(l)}}^\top \\ &= \underbrace{\sigma_{3D}^2 \sum_{i=1}^{N} \frac{\partial \Phi}{\partial M_i} . \frac{\partial \Phi}{\partial M_i}^\top}_{\Gamma_{MM}} + \underbrace{\sigma_{2D}^2 \sum_{l=1}^{M} \sum_{i=1}^{N} \frac{\partial \Phi}{\partial \tilde{m}_i^{(l)}} . \frac{\partial \Phi}{\partial \tilde{m}_i^{(l)}}^\top}_{\Gamma_{mm}} \end{aligned}$$

We may finally write the covariance of T:

$$\boxed{\Sigma_{TT} = H^{-1} . (\Gamma_{MM} + \Gamma_{mm}) . H^{-T}} \tag{15}$$

with the following simplifications in the expressions:

$$\Gamma_{mm} = \frac{1}{\sigma_{2D}^2}\sum_{i=1}^{N}\sum_{l=1}^{M}\xi_i^l \cdot \frac{{D_i}^\top \cdot \left(Q^{(l)} - m_i^{(l)} \cdot C_{(l)}^\top\right)^\top \cdot \left(Q^{(l)} - m_i^{(l)} \cdot C_{(l)}^\top\right) \cdot D_i}{(1+<C_{(l)}|T \star M_i>)^2} = \sum_{i=1}^{N} {D_i}^\top \cdot L_i \cdot D_i$$

$$\Gamma_{MM} = \sigma_{3D}^2 \sum_{i=1}^{N} {D_i}^\top \cdot K_i \cdot K_i \cdot D_i$$

$$H = \sum_{i=1}^{N} {D_i}^\top \cdot K_i \cdot D_i$$

3.3 Theoretical Uncertainty of the EPPC

We now turn to the uncertainty prediction of the EPPC given in Eq. 6. Dropping the constant term, this criterion is:

$$C(T, M_1, \dots M_N) = \sum_{i=1}^{N} \frac{\| \tilde{M}_i - M_i \|^2}{2 \cdot {\sigma_{3D}}^2} + \sum_{l=1}^{M}\sum_{i=1}^{N} \xi_i^l \cdot \frac{\| \tilde{m}_i^{(l)} - m_i^{(l)} \|^2}{2 \cdot {\sigma_{2D}}^2}$$

where $m_i^{(l)} = P^{(l)}(T \star M_i)$ is define as in the previous section. Thus, its derivatives given in Eq. 9 and 10 are still valid.

Implicit equation of the optimum This time, the minimization is performed not only on the transformation parameters T but also on the perfect 3D coordinates M_i. Thus, we have at the optimum:

$$\Phi_T = \left(\frac{\partial C}{\partial T}\right)^\top = 0 \qquad \text{and} \qquad \Phi_{M_i} = \left(\frac{\partial C}{\partial M_i}\right)^\top = 0$$

Since the first term of the criterion does not directly depends upon the transformation, we may reuse directly the results of the last section to obtain Φ_T:

$$\Phi_T = \frac{1}{\sigma_{2D}^2}\sum_{l=1}^{M}\sum_{i=1}^{N} \xi_i^l \cdot {D_i}^\top \frac{(Q^{(l)} - m_i^{(l)}.C_{(l)}^\top)^\top (m_i^{(l)} - \tilde{m_i}^{(l)})}{1+ < C_{(l)}|T \star M_i >} \tag{16}$$

Then, a quick step clarifies Φ_{M_i}:

$$\begin{aligned}\Phi_{M_i} &= \left(\frac{(M_i - \tilde{M}_i)^\top}{\sigma_{3D}^2} + \sum_{l=1}^{M} \xi_i^l \cdot \frac{\left(m_i^{(l)} - \tilde{m}_i^{(l)}\right)^\top}{\sigma_{2D}^2} \cdot \frac{\partial m_i^{(l)}}{\partial M_i} \right)^\top \\ &= \frac{(M_i - \tilde{M}_i)}{\sigma_{3D}^2} + \sum_{l=1}^{M} \xi_i^l \cdot \frac{R^\top . \left(Q^{(l)} - m_i^{(l)}.{C^{(l)}}^\top\right)^\top . \left(m_i^{(l)} - \tilde{m}_i^{(l)}\right)}{\sigma_{2D}^2 . (1+ < C^{(l)}|T \star M_i >)} \end{aligned} \tag{17}$$

Variability of the optimal transformation alone In our particular application, the exact 3D points M_i are only auxiliary variables and not parameters of interest: we are looking for the covariance $\Sigma_{TT} = E\left((T-\bar{T})(T-\bar{T})^\top\right) = E(\delta T.\delta T^\top)$ of the transformation only, and not the joint covariance of all parameters. Let us regroup the coordinates of all the exact 3D points into M, so that the parameter vector and the implicit function defining the optimum become $p = \begin{bmatrix} T \\ M \end{bmatrix}$ and $\Phi(p,\chi) = \begin{bmatrix} \Phi_T \\ \Phi_M \end{bmatrix} = 0$. In this section, the data vector χ gathers the coordinates of the observed 3D and 2D points $\tilde{M}_i$ and $\tilde{m}_i^{(l)}$. When the data changes from χ to $\chi + \delta\chi$, the optimal parameters changes from p (where $\Phi(p,\chi) = 0$) to $p + \delta p$ such that, up to the second order, we have:

$$\Phi(p+\delta p, \chi+\delta\chi) = \frac{\partial\Phi}{\partial p}.\delta p + \frac{\partial\Phi}{\partial\chi}.\delta\chi = 0$$

This means that:

$$\delta p = \begin{bmatrix} \delta T \\ \delta M \end{bmatrix} = -\left(\frac{\partial\Phi}{\partial p}\right)^{-1} \cdot \frac{\partial\Phi}{\partial\chi}.\delta\chi = -\begin{bmatrix} \frac{\partial\Phi_T}{\partial T} & \frac{\partial\Phi_T}{\partial M} \\ \frac{\partial\Phi_M}{\partial T} & \frac{\partial\Phi_M}{\partial M} \end{bmatrix}^{(-1)} \cdot \begin{bmatrix} \frac{\partial\Phi_T}{\partial\chi} \\ \frac{\partial\Phi_M}{\partial\chi} \end{bmatrix}.\delta\chi$$

At this step, we could calculate the analytical forms of $\frac{\partial\Phi}{\partial p}$ and $\frac{\partial\Phi}{\partial\chi}$, use a numerical method to inverse the matrix, compute $\Sigma_{pp} = E\left(\delta p.\delta p^\top\right) = \begin{bmatrix} \Sigma_{TT} & \Sigma_{TM} \\ \Sigma_{MT} & \Sigma_{MM} \end{bmatrix}$ and finally extract Σ_{TT} from this big matrix. Nevertheless, it si more judicious to push the computations one step further in order to solve analytically for δT since the final formula will be easier to understand and to implement. Moreover, we avoid the tiresome and useless computations of δM because we focus only on the first part parameters' uncertainties, i.e. the transformation.

To analytically express $\frac{\partial\Phi}{\partial p}^{-1}$, we make a block inversion with $A = \frac{\partial\Phi_T}{\partial T}$, $B = \frac{\partial\Phi_T}{\partial M}$ and $C = \frac{\partial\Phi_M}{\partial M}$:

$$\begin{bmatrix} A & B \\ B^\top & C \end{bmatrix} \cdot \begin{bmatrix} D & E \\ E^\top & F \end{bmatrix} = Id \quad \Longleftrightarrow \quad \begin{cases} D &= & (A - B\cdot C^{-1}\cdot B^\top)^{-1} \\ F &= & (C - B^\top\cdot A^{-1}\cdot B)^{-1} \\ E &= & -A^{-1}\cdot B(C - B^\top\cdot A^{-1}\cdot B)^{-1} \\ &= & -(A - B\cdot C^{-1}\cdot B^\top)^{-1}\cdot B\cdot C^{-1} \end{cases}$$

Thus, we obtain the analytical expression of δT:

$$\delta T = \left(D\cdot\frac{\partial\Phi_T}{\partial\chi} + E\cdot\frac{\partial\Phi_M}{\partial\chi}\right)\cdot\delta\chi = (A - B\cdot C^{-1}\cdot B^\top)^{-1}\left[\frac{\partial\Phi_T}{\partial\chi} - B\cdot C^{-1}\cdot\frac{\partial\Phi_M}{\partial\chi}\right]\cdot\delta\chi \quad (18)$$

Partial derivatives of Φ Thus, since $\frac{\partial \Phi_{M_i}}{\partial M_j} = 0$ for $i \neq j$, we are left with the computation of

$$A = H_T = \frac{\partial \Phi_T}{\partial T}, \quad B = \frac{\partial \Phi_T}{\partial M} = \left[\frac{\partial \Phi_T}{\partial M_1}, \cdots, \frac{\partial \Phi_T}{\partial M_N}\right] \quad \text{and} \quad C = H_M = \begin{bmatrix} \frac{\partial \Phi_{M_1}}{\partial M_1} & & 0 \\ & \ddots & \\ 0 & & \frac{\partial \Phi_{M_N}}{\partial M_N} \end{bmatrix}$$

in addition to:

$$\frac{\partial \Phi_T}{\partial \chi} = \left[\dots \frac{\partial \Phi_T}{\partial \tilde{M}_i} \dots \frac{\partial \Phi_T}{\partial \tilde{m}_i^{(l)}} \dots\right] \qquad \text{and} \qquad \frac{\partial \Phi_{M_j}}{\partial \chi} = \left[\dots \frac{\partial \Phi_{M_j}}{\partial \tilde{M}_i} \dots \frac{\partial \Phi_{M_j}}{\partial \tilde{m}_i^{(l)}} \dots\right]$$

Thank to the result of the previous section (Eq. 12, 13 and 14), we already know that:

$$\begin{aligned} H_T = \frac{\partial \Phi_T}{\partial T} &= \sum_{i=1}^{N} D_i^\top \cdot K_i \cdot D_i \\ \frac{\partial \Phi_T}{\partial \tilde{m_i}^{(l)}} &= -\xi_i^l \cdot \frac{1}{\sigma_{2D}^2} \frac{{D_i}^\top \cdot \left(Q^{(l)} - m_i^{(l)} \cdot C_{(l)}^\top\right)^\top}{1 + < C_{(l)} | T \star M_i >} \\ B_i = \frac{\partial \Phi_T}{\partial M_i} &= D_i^\top \cdot K_i \cdot R \end{aligned}$$

The other derivatives are similarly computed:

$$\begin{aligned} \frac{\partial \Phi_T}{\partial \tilde{M}_i} &= 0 \\ H_{M_i} = \frac{\partial \Phi_{M_i}}{\partial M_i} &= \frac{Id_3}{\sigma_{3D}^2} + R^t \cdot K_i \cdot R \\ \frac{\partial \Phi_{M_i}}{\partial \tilde{M}_j} &= -\delta_{ij} . \frac{Id_3}{\sigma_{3D}^2} \\ \frac{\partial \Phi_{M_i}}{\partial \tilde{m}_j^{(l)}} &= -\delta_{ij} \cdot \xi_j^l \cdot \frac{R^\top \cdot (Q - m_j^{(l)} \cdot C_{(l)}^\top)^\top}{\sigma_{2D}^2 (1 + < C_{(l)} | T \star M_i >)} \end{aligned}$$

Covariance of the transformation Let us write:

$$\begin{aligned} H &= A - B \cdot C^{-1} \cdot B^\top = H_T - B \cdot H_M^{-1} \cdot B^\top \\ \Gamma &= \left[\frac{\partial \Phi_T}{\partial \chi} - B \cdot H_M^{-1} \cdot \frac{\partial \Phi_M}{\partial \chi}\right] \Sigma_{\chi\chi} \left[\frac{\partial \Phi_T}{\partial \chi} - B \cdot H_M^{-1} \cdot \frac{\partial \Phi_M}{\partial \chi}\right]^\top \end{aligned}$$

Using Eq. 18, the covariance of the transformation may be written:

$$\Sigma_{TT} = E\left(\delta T \cdot \delta T^{\top}\right) = H^{(-1)} \cdot \Gamma \cdot H^{(-1)}$$

To compute Γ, let us first consider the components relatively to $\tilde{M}_i$ and $\tilde{m}_i^{(l)}$ of χ in the matrix $\left[\frac{\partial \Phi_T}{\partial \chi} - B \cdot C^{-1}.\frac{\partial \Phi_M}{\partial \chi}\right]$:

$$\left[\frac{\partial \Phi_T}{\partial \chi} - B \cdot C^{-1}.\frac{\partial \Phi_M}{\partial \chi}\right]_{\tilde{M}_i} = \underbrace{\frac{\partial \Phi_T}{\partial \tilde{M}_i}}_{0} - \sum_{k=1}^{N} B_k \cdot H_{M_k}^{-1}.\frac{\partial \Phi_{M_k}}{\partial \tilde{M}_i} = \frac{B_i \cdot H_{M_i}^{-1}}{\sigma_{3D}^2} \quad (19)$$

$$\begin{aligned}\left[\frac{\partial \Phi_T}{\partial \chi} - B \cdot C^{-1} \cdot \frac{\partial \Phi_M}{\partial \chi}\right]_{\tilde{m}_i^{(l)}} &= \frac{\partial \Phi_T}{\partial \tilde{m}_i^{(l)}} - \sum_{k=1}^{N} B_k \cdot H_{M_k}^{-1} \cdot \frac{\partial \Phi_{M_k}}{\partial \tilde{m}_i^{(l)}} \\ &= \xi_i^l \cdot (B_i \cdot H_{M_i}^{-1} \cdot R^{\top} - {D_i}^{\top}) \frac{(Q - m_i^{(l)} \cdot C_{(l)}^{\top})^{\top}}{\sigma_{2D}^2 . (1 + < C_{(l)} | T \star M_i >)} \quad (20)\end{aligned}$$

The 3D points $\tilde{M}_i$ and the 2D points $\tilde{m}_i^{(l)}$ being independent, $\Sigma_{\chi\chi}$ is a diagonal block matrix. Moreover, we assume an isotropic and identical noise on all 3D points and on all 2D points (i.e. $\Sigma_{M_i} = \sigma_{3D}^2 \cdot \mathrm{Id}_3$ and $\Sigma_{\tilde{m}_i^{(l)}} = \sigma_{2D}^2 \cdot \mathrm{Id}_2$).Using Eq. (19) and (20) We can then write Γ as:

$$\Gamma = \sum_{i=1}^{N} \frac{B_i \cdot H_{M_i}^{-1} H_{M_i}^{-1} \cdot B_i^{\top}}{\sigma_{3D}^2} + \sum_{i=1}^{N} \sum_{l=1}^{M} \xi_i^l \cdot (B_i \cdot H_{M_i}^{-1} \cdot R^{\top} - {D_i}^{\top}) \cdot L_i^{(l)} \cdot (B_i \cdot H_{M_i}^{-1} \cdot R^{\top} - {D_i}^{\top})^{\top}$$

To further simplify this expression, we should note that

$$B_i \cdot H_{M_i}^{-1} \cdot R^{\top} = {D_i}^{\top} \cdot \left(\frac{K_i^{-1}}{\sigma_{3D}^2} + Id_3\right)^{-1}$$

This expression will be used to simplify the first term of Γ. For the second term, we use the inversion lemma $(A+B)^{-1} = A^{-1} - A^{-1}(B^{-1} + A^{-1})^{-1}A^{-1}$ to obtain:

$$B_i \cdot H_{M_i}^{-1} \cdot R^{\top} = {D_i}^{\top} \cdot \left(Id_3 - \left(\sigma_{3D}^2 Id_3 + K_i^{-1}\right)^{-1} \cdot K_i^{-1}\right)$$

This leads to:

$$\Gamma = \sum_{i=1}^{N} \frac{{D_i}^{\top}}{\sigma_{3D}^2} \cdot \left(\frac{K_i^{-1}}{\sigma_{3D}^2} + Id\right)^{-2} \cdot D_i + \sum_{i=1}^{N} {D_i}^{\top} (\sigma_{3D}^2 Id + K_i^{-1})^{-1} \cdot K_i^{-1} \cdot L_i \cdot K_i^{-1} \cdot (\sigma_{3D}^2 Id + K_i^{-1})^{-1} \cdot D_i$$

and finally after further simplifications to:

$$\Gamma = \sum_{i=1}^{N} {D_i}^\top (\sigma_{3D}^2 Id + K_i^{-1})^{-1} (\sigma_{3D}^2 Id + K_i^{-1} \cdot L_i \cdot K_i^{-1}) (\sigma_{3D}^2 Id + K_i^{-1})^{-1} \cdot D_i \quad (21)$$

We are now left with the simplification of H. Thanks to the previously established derivatives, we easily get:

$$H = H_T - \sum_{i=1}^{N} B_i . H_{M_i}^{-1} . B_i^\top = \sum_{i=1}^{N} {D_i}^\top \cdot K_i \cdot D_i - \sum_{i=1}^{N} {D_i}^\top \cdot \left[K_i \cdot \left(\frac{Id}{\sigma_{3D}^2} + K_i \right)^{-1} \cdot K_i \right] \cdot D_i$$

Using once again the inversion lemma, we eventually obtain:

$$H = \sum_{i=1}^{N} {D_i}^\top \cdot \left(\sigma_{3D}^2 \cdot Id + K_i^{-1} \right)^{-1} \cdot D_i \quad (22)$$

3.4 Algorithmic summary

We finally obtained a quite simple expressions of Σ_{TT} for both criteria relying on a few intermediate variables. We summarize in this section the computation steps needed to implement its computation.

Common variables

$$\begin{aligned}
D_i &= \tfrac{\partial (T \star M_i)}{\partial T} \\
m_i^{(l)} &= P^{(l)}(T \star M_i) = \frac{Q^{(l)} \cdot (T \star M_i) + b^{(l)}}{1 + C_{(l)}^\top \cdot (T \star M_i)} \\
L_i &= \textstyle\sum_{l=1}^{M} \xi_i^l \cdot \frac{(Q - m_i^{(l)} \cdot C_{(l)}^\top)^\top \cdot (Q - m_i^{(l)} \cdot C_{(l)}^\top)}{\sigma_{2D}^2 \cdot (1 + <C_{(l)} | T \star M_i>)^2} \\
K_i &= L_i - \textstyle\sum_{l=1}^{M} \xi_i^l \cdot \frac{C_{(l)} \cdot (m_i^{(l)} - \tilde{m}_i{}^{(l)})^\top \cdot (Q^{(l)} - m_i^{(l)} \cdot C_{(l)}^\top) + (Q^{(l)} - m_i^{(l)} \cdot C_{(l)}^\top)^\top \cdot (m_i^{(l)} - \tilde{m}_i{}^{(l)}) \cdot C_{(l)}^\top}{\sigma_{2D}^2 \cdot (1 + <C_{(l)} | T \star M_i>)^2}
\end{aligned}$$

Intermediate variables for the SPPC For this criterion, we remind that the observed 3D data are considered as exact, so that $M_i = \tilde{M}_i$ is fed into the above variables.

$$\begin{aligned} \Gamma &= \sum_{i=1}^{N} {D_i}^\top \left(\sigma_{3D}^2 \cdot K_i \cdot K_i + L_i\right) \cdot D_i \\ H &= \sum_{i=1}^{N} {D_i}^\top \cdot K_i \cdot D_i \end{aligned}$$

Intermediate variables for the EPPC For this criterion, the exact 3D point positions M_i are determined by the minimization of the criterion, along with the optimal transformation.

$$\begin{aligned} \Gamma &= \sum_{i=1}^{N} {D_i}^\top \left(\sigma_{3D}^2 \cdot Id + K_i^{-1}\right)^{-1} \cdot \left(\sigma_{3D}^2 \cdot Id + K_i^{-1} \cdot L_i \cdot K_i^{-1}\right) \cdot \left(\sigma_{3D}^2 \cdot Id + K_i^{-1}\right)^{-1} \cdot D_i \\ H &= \sum_{i=1}^{N} {D_i}^\top \cdot \left(\sigma_{3D}^2 \cdot Id + K_i^{-1}\right)^{-1} \cdot D_i \end{aligned}$$

Final uncertainty estimation For both criteria, the covariance matrix on the transformation is finally computed as:

$$\Sigma_{TT} = H^{-1} \cdot \Gamma \cdot H^{-1}$$

To obtain the final covariance matrix on a target point C_i after registration, we simply have to propagate the uncertainty through the transformation action:

$$\Sigma_{T \star C_i} = \frac{\partial (T \star C_i)}{\partial T} \cdot \Sigma_{TT} \cdot \frac{\partial (T \star C_i)}{\partial T}^\top$$

Equivalence of the two criteria uncertainties for perfect 3D data For the limit case where $\sigma_{3D} = 0$, we know that the two criteria are equivalent. We can check that their uncertainty is also equal. Indeed, we have for both criteria the following simplifications:

$$\begin{aligned} \Gamma(\sigma_{3D} = 0) &= \sum_{i=1}^{N} {D_i}^\top \cdot K_i \cdot (K_i^{-1} \cdot L_i \cdot K_i^{-1}) \cdot K_i \cdot D_i = \sum_{i=1}^{N} {D_i}^\top \cdot L_i \cdot D_i \\ H(\sigma_{3D} = 0) &= \sum_{i=1}^{N} {D_i}^\top \cdot K_i \cdot D_i \end{aligned}$$

4 Performances Evaluation and Comparison of Criteria

This section is devoted to an evaluation and comparison of SPPC and EPPC. We first evaluate the criteria on synthetic data, which shows that EPPC is much more robust than SPPC and that it provides a gain in accuracy of up to 20% depending on the geometric configuration of the cameras. Then, we turn to an accuracy assessment using real images from a phantom of the abdomen. Since the "skin" of the phantom (on which are stuck the markers) is flexible, these data are very close to real clinical conditions. This study proves that EPPC brings in this case an accuracy improvement of up to 9% and that a mean accuracy of about 2 mm can be reached.

4.1 Evaluation with Synthetic Data

The goal is to assess the effectiveness of the SPPC and EPPC criteria in terms of computing cost, accuracy and robustness. This study is especially important important in view of the constraints of our application: we aims at performing a real-time registration of the internal organs on the video images with an accuracy better than 5 mm, at a refreshment rate of 25 Hz. Thus, we have 0.04 seconds to provide the registration, the images data transmission and the monitoring of the reconstructed models. Thus, it is crucial to assess the trade-off to make between the additional accuracy provided by the EPPC and the time cost that this gain implies. Of course, the speed and accuracy performances appraisals need to be related to the data feature used and to the parameters of both algorithms which are almost identical (the EPPC only needs one additional parameter to decide when the alternated minimization has to finish), and whose value are identical by default.

4.1.1 Experimental Setup

Experiments are realized with two synthetic cameras focusing the same point at an angle of 45 degrees and jointly calibrated in the same reference frame R_f. We use 15 points N_i in a box of about 15×15×15 mm, whose 3D coordinates are known in R_f, and whose projection m_i in the video image were accurately computed (pixel units). Both camera are 1300 mm away from the center of mass of the point set.

To simulate the fact that the 3D points are obtained in an another frame than R_f (in our case the scanner frame), we apply a known rigid transformation T^{-1} (parameterized by the rotation and translation vectors) to the N_i and the resulting points are named M_i. Then, we model the human localization error with a centered Gaussian noise of standard deviation 2, that corrupt the x, y, z coordinates of the

M_i (in mm for the voxel images) and the u, v coordinates of the m_i (in pixels for the video images). To follow the previous naming conventions, the disturbed points will be written $\tilde{M}_i$ and $\tilde{m}_i$. Therefore, the optimization step provides an estimation $\hat{T}$ of the transformation T. The optimization procedure used for the transformation and for the 3D points coordinates relies on the Powell algorithm [Pow64, PFTV92] with a minimization tolerance of 10^{-4}.

4.1.2 Performances Measures

The registration error is evaluated using nine carefully chosen control points C_i defined in scanner reference frame R_f, by computing the RMS error between the $\hat{T} \star T^{-1} \star C_i$ and the C_i:

$$\text{RMS Error } = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \| \hat{T} \star T^{-1} \star C_i - C_i \|^2}$$

Please, notice that this RMS error corresponds in fact to the *Target Registration Error* (TRE) commonly employed in medical applications to evaluate the registration accuracy, and not to a *Fiducial Localization Error* (FLE), that would only reflect the noise on the data points.

The default values used for generating the data are: 15 points with a 2D (resp. 3D) standard deviation of 2 (resp. 2 mm), an angle between the cameras of 45°, and a seek rotation (resp. translation) of 1.1 rad = 63 deg (resp. 60 mm).

Each table presented in this section shows the comparative values of the computing time and the accuracy of both criteria with respect to a single parameter, other parameters being fixed to the previous default values. For each value of the varying parameter, we give the mean computation time and RMS error over 10 000 registrations experiments. To further detail the variability of the accuracy performances, we also present the standard deviation of the measured RMS and its maximum value.

As an additional indication for the comparison of the two criteria, we give the relative error and the number of time that the EPPC gives a better accuracy among the 10 000 registrations. In order to give the symmetry property to the relative error measurements, we compute in fact the exponential of the log-mean of the RMS ratio:

$$\text{Relative error} = \exp\left(\frac{1}{10000} \cdot \sum_{1}^{10000} \log\left(\frac{RMS_{\text{SPPC}}}{RMS_{\text{EPPC}}}\right)\right)$$

Thus, a value above one indicates that the EPPC is relatively more accurate than the SPPC.

4.1.3 Computations Time and Accuracy

Choice of stopping value d_T for the EPPC alternated minimization Table 1 assess the influence on the EPPC accuracy of the stopping threshold d_T on the transformation evolution (defined in Sec. 2.3.2). In this section we abusively redefine d_T so that:

$$d^r_{max} = d_T \cdot \Pi \quad \text{and} \quad d^t_{max} = d_T \cdot dim$$

where dim is chosen as the size of the object to register.

We may notice that the RMS error and the relative error do not significantly change below 10^{-4}, and become slightly worst above. On the other hand, the computation time is very sensitive to d_T: it changes from 0.026 sec. to 0.961 sec. for a d_T value 1 million time smaller and an accuracy result which do not vary. Thus, we can conclude that choosing $d_T = 10^{-4}$ is a very good tradeoff between the accuracy and the computation time of the EPPC.

d_T value of the EPPC		10^{-1}	10^{-3}	10^{-4}	10^{-5}	10^{-7}	10^{-10}
Computations time	SPPC	0.0024s	0.0026s	0.0026s	0.0024s	0.0025s	0.0024s
	EPPC	0.011s	0.014s	0.026s	0.073s	0.297s	0.961s
Mean RMS error in mm	SPPC	2.15	2.14	2.14	2.14	2.13	2.17
	EPPC	2.07	2.00	1.96	1.95	1.94	1.97
Std. dev. of the mean RMS error	SPPC	0.83	0.83	0.83	0.84	0.84	0.86
	EPPC	0.79	0.77	0.74	0.74	0.73	0.75
Maximal RMS error in mm	SPPC	6.63	6.44	6.56	6.32	6.42	6.60
	EPPC	6.21	5.26	5.56	5.68	5.62	5.77
Mean predicted RMS error in mm	SPPC	2.19	2.19	2.19	2.19	2.19	2.19
	EPPC	2.01	2.01	2.01	2.01	2.01	2.01
Relative error		1.038	1.062	1.084	1.088	1.076	1.091
Number of success		7099	6937	6471	6353	6295	6302

Table 1: Computations time and accuracy w.r.t. the EPPC stopping value.

Influence of the minimization tolerance The results of Table 2 show us that the computation time and the accuracy of the SPPC do not significantly depend on the minimization tolerance (Powell's algorithm) as long as it is inferior to 10^{-2}. For values 10 times superior, computation time are lower (0.0019 sec. against 0.0022 sec.), but the accuracy degrades drastically (a factor 2 on the RMS error!). The EPPC seems to behave like the SPPC even for large value (i.e. 10^{-1}): the computation times and the RMS error are almost the same for the 5 tolerance values used. We

could have assumed the contrary as the EPPC corresponds to an iteration of the SPPC. In fact, the EPPC will compensate with more iterations. Each iteration will be faster but their number will be more important at the end. That's why it looks like if it does not depend on the tolerance parameter.

In our experimental setup, we adopted a value of 10^{-4}. We will see in the robustness study (Section 4.1.4) that this value was good enough to carry out correctly the minimization for that case. In fact, for a random seeked transformation, it is necessary to decrease the tolerance value to avoid bad convergence.

Minimization tolerance value		10^{-5}	10^{-4}	10^{-3}	10^{-2}	10^{-1}
Computations time	SPPC	0.0027s	0.0025s	0.0024s	0.0022s	0.0019s
	EPPC	0.033s	0.025s	0.024s	0.022s	0.029s
Mean RMS error in mm	SPPC	2.14	2.14	2.14	2.15	4.11
	EPPC	1.94	1.96	1.99	2.00	2.04
Standard deviation of the mean RMS error	SPPC	0.84	0.84	0.83	0.83	2.16
	EPPC	0.74	0.75	0.76	0.76	0.78
Maximal RMS error in mm	SPPC	6.01	6.54	6.18	6.27	11.27
	EPPC	5.57	6.43	6.00	5.31	5.86
Mean predicted RMS error in mm	SPPC	2.19	2.19	2.19	2.19	2.19
	EPPC	2.01	2.01	2.01	2.01	2.01
Relative error		1.095	1.085	1.066	1.070	1.735
Number of success		6472	6560	7085	7166	8754

Table 2: Computations time and accuracy w.r.t. the minimization tolerance.

Influence of the noise amplitude Table 3 shows that the noise amplitude has a quite linear influence on the accuracy. Indeed, the RMS error of both criteria is roughly equal (in this case) to the standard deviation of the 2D and 3D noises, even if the EPPC is always slightly better.

On the other hand, when the 3D noise increase w.r.t. the 2D noise, Table 4 shows that the relative error increase as well (1.09 relative error for a unit 3D to 2D standard deviation ratio against 1.15 for a ratio of 2). This means that the EPPC is more and more accurate, which was expected since this criterion better captures the noisy nature of information on the 3D data. This gain in accuracy is nevertheless compensated by a much heavier computation cost (0.025 sec. against 0.057 sec. in the last example).

Noise std. dev. (2D/3D)		0.2 / 0.2	1 / 1	2 / 2	3 / 3	4 / 4	10 / 10
Computation time	SPPC	0.0027s	0.0026s	0.0026s	0.0025s	0.0026s	0.0024s
	EPPC	0.014s	0.025s	0.025s	0.025s	0.025s	0.050s
Mean RMS error in mm	SPPC	0.21	1.08	2.16	3.21	4.30	11.09
	EPPC	0.20	0.98	1.97	2.92	3.91	9.88
Std. dev. of the mean RMS error	SPPC	0.08	0.42	0.84	1.24	1.68	4.36
	EPPC	0.07	0.38	0.75	1.10	1.49	3.77
Maximal RMS error in mm	SPPC	0.56	3.10	6.38	9.16	12.03	30.35
	EPPC	0.55	2.72	5.82	7.49	12.02	27.14
Mean predicted RMS error in mm	SPPC	0.27	1.12	2.19	3.25	4.30	10.31
	EPPC	0.26	1.04	2.01	2.98	3.95	9.67
Relative error		1.062	1.083	1.092	1.089	1.088	1.109
Number of success		7019	6650	6710	6572	6619	6577

Table 3: Computations time and accuracy w.r.t. of the noise corrupting the data.

Ratio of noises std. dev.: $\frac{\sigma_{3D}}{\sigma_{2D}}$		8	4	2	1	0.5	0.25
Range of σ_{3D} in which the relative error is constant		2 - 6	3 - 8	1 - 7	1 - 8	0.5 - 4	0.1 - 2
Computations time (sec.)	SPPC	0.0024	0.0024	0.0023	0.0024	0.0023	0.0024
	EPPC	0.055	0.119	0.057	0.025	0.024	0.021
Relative error		1.08	1.17	1.15	1.09	1.03	1.006
Number of success		7953	7378	6982	6652	5928	5432

Table 4: Computations time and relative error w.r.t. 3D/2D noise ratio on the data.

Surprisingly, Table 4 also shows that there is a limit to this behavior: when the 3D noise becomes too large ($\sigma_{3D}^2/\sigma_{2D}^2$ ratio equal to 8), the relative error and the computation times decrease instead of increasing.

Influence of the angle between the cameras Table 5 presents the results of experiments with an angle between the cameras ranging from 10 to 90 degrees. We first notice that the RMS error decreases as the angle reaches 90°. This was predictable since the depth of points is better constrained as the angle between the cameras widen. However, the EPPC is all the better as the angle decrease, up to a significant relative error of 1.188 for an angle of 10°. This evolution is coherent

as the amount of 3D information depends on the angle between the camera. When this information is maximal (90°), the EPPC brings no more improvement w.r.t. the SPPC (relative error = 1.03).

Angle between the cameras		90°	60°	45°	30°	10°
Computation time	SPPC	0.0025s	0.0034s	0.0026s	0.0025s	0.0031s
	EPPC	0.015s	0.021s	0.025s	0.028s	0.022s
Mean RMS error in mm	SPPC	1.76	1.94	2.14	2.46	4.12
	EPPC	1.75	1.83	1.94	2.22	3.40
Standard deviation of the mean RMS error	SPPC	0.67	0.74	0.84	0.96	1.76
	EPPC	0.66	0.69	0.74	0.85	1.39
Maximal RMS error in mm	SPPC	4.89	6.64	5.92	6.50	12.40
	EPPC	4.81	5.78	5.18	6.06	9.87
Mean predicted RMS error in mm	SPPC	1.82	1.98	2.19	2.50	4.11
	EPPC	1.81	1.89	2.01	2.27	3.54
Relative error		1.008	1.051	1.089	1.114	1.188
Number of success		5586	6183	6584	6765	6772

Table 5: Computations time and accuracy w.r.t. the angle between the cameras.

Influence of the number of points We end our comparative study with Table 6 that investigates the number of points used for the registration. We first notice that the computation times are almost proportional to the number of points. This was foreseeable since the computational complexity is linear in the number of data points. The second observation is that the RMS seems to be inversely proportional to the square root of the number of points's variation (multiplying 2 times the number of points decrease the RMS error by a factor $\sqrt{2}$), which is also in accordance with the standard accuracy improvements in statistics.

4.1.4 Robustness Evaluation

This part deals with a comparison of the ratio of wrong convergence (RWC) of both criteria w.r.t. the initial data. We particularly focus on the influence of the seeked and initial transformation. The experimental conditions are globally the same as in Section 4. We just choose randomly the seeked transformation and/or the initialization. The rotation of the transformation is parameterized with a rotation vector whose norm (resp. direction) is chosen uniformly in the range $]-\pi;\pi]$ (respect. in the unit sphere). Each coordinate of the translation is chosen uniformly in the

Number of points		30	15	8	4
Computation time	SPPC	0.0042s	0.0026s	0.0016s	0.0011s
	EPPC	0.067s	0.025s	0.011s	0.006s
Mean RMS error in mm	SPPC	1.55	2.16	2.73	4.45
	EPPC	1.42	1.97	2.51	3.94
Standard deviation of the mean RMS error	SPPC	0.61	0.83	1.05	1.78
	EPPC	0.55	0.74	0.94	1.53
Maximal RMS error in mm	SPPC	4.63	6.40	7.80	12.21
	EPPC	3.84	5.80	7.15	10.64
Mean predicted RMS error in mm	SPPC	1.59	2.19	2.79	4.47
	EPPC	1.46	2.01	2.57	3.98
Relative error		1.080	1.090	1.079	1.117
Number of success		6644	6607	6373	6538

Table 6: Computations time and accuracy w.r.t. the number of points.

range $]-750;750]$. This choice should not be far from the real experience in which the cameras are positioned just on the side of the scanner. Then, the translation between the two references frame is around 1 meter.

We compute the same performance measures as in the last section and we add the RWC for each criterion. The RMS error values corresponding to a wrong convergence are obviously not taken into account for the computation of the mean RMS error.

To decide if the algorithm has correctly converged for each registration, we compare the RMS error found with the maximum RMS error computed in the experience of the last section that presents the same initial conditions. Indeed, it is very unlikely that the algorithm converges to the global minimum with a RMS error above the maximum RMS computed before since the initial data are identical to those of the last section (except for the seeked transformation T).

A better Tolerance Minimization Value In the last section, we use a Powell tolerance value of 10^{-4} that was sufficient to ensure a good convergence for the chosen transformation. We have experienced that this value was not small enough to carry out a good convergence when the transformation and the initialization were chosen randomly. Table 7 presents the influence of this value on both criteria robustness.

A brief consideration of the results about the SPPC shows clearly that we should use a tolerance value of 10^{-14} if we want to avoid most of the wrong convergences (a smaller tolerance value does not add any accuracy or robustness because we reach

Powell Tolerance Value		10^{-4}	10^{-5}	10^{-6}	10^{-7}	10^{-8}	10^{-10}	10^{-14}
Computation time (sec.)	SPPC	0.0078	0.0101	0.0117	0.0141	0.0162	0.0193	0.0244
	EPPC	0.478	0.229	0.383	0.396	0.536	0.607	0.840
Mean RMS error in mm	SPPC	3.03	3.02	2.66	2.55	2.43	2.49	2.40
	EPPC	2.61	2.14	1.99	1.99	1.98	1.95	1.94
Std. dev. of the mean RMS error	SPPC	1.53	1.45	1.27	1.26	1.12	1.20	1.11
	EPPC	1.07	0.85	0.73	0.76	0.76	0.73	0.75
Maximal RMS error in mm	SPPC	6.71	6.66	6.71	6.67	6.55	6.61	6.66
	EPPC	5.92	5.96	5.47	5.65	5.20	4.52	4.49
Ratio of wrong convergence	SPPC	84.1%	78.3%	70.5%	63.0%	56.9%	44.9%	29.0%
	EPPC	10.5%	3.0%	1.6%	0.4%	0.6%	0.4%	0.4%
Relative error		1.214	1.298	1.237	1.198	1.182	1.175	1.163
Number of success		849	921	900	843	836	805	755

Table 7: Influence of the Powell tolerance value on the criteria robustness.

the limits of the computing numerical accuracy). Considering now the EPPC, we see firstly that it is less sensitive to the tolerance value than the SPPC (its range of wrong convergence is between 0.4% and 10.5%, whereas it is between 29% and 84% for the short criterion). Secondly, it does not seem very useful to choose a value smaller than 10^{-7} to reach the whole efficiency of the complete criterion. Therefore, we decide to choose for the following robustness experiments to fix the tolerance to 10^{-14} (resp. 10^{-7}) for the short (resp. complete) criterion.

Robustness Results Now that we found equitable tolerance values to compare both criteria, we present in this section three different kind of experiments to assess the robustness of the criteria. The first three tables corresponds to the case where T is randomly chosen while initialize to the identity. For most registration problems, this represents the most likely real experience: when we have no prior information about the transformation, it is reasonable to initialize the algorithm with a null rotation and translation. Then, exchanging the role of the seeked and initial transformation, we observe in Table 11 that the minimization algorithm does not behave symmetrically. This points out that the minimization method might be biased toward the identity. The last table offers interesting results on the case where both the initialization and the seeked transformation are randomly chosen. More clearly, it corresponds to the occurrence of a very bad initialization.

Random T with Identity Initialization: Table 8 shows the incidence of the noise amplitude on the robustness criteria. The computations time does not seem to be affected for both criteria, whereas the ratio of wrong convergence (RWC) decrease slightly from 21.9% to 18.4% for the SPPC when the noise increase. For the EPPC, the RWC might not be interpretable as the value are very close to 0.

Amplitude of the noise 3D/2D		0.2/0.2	1/1	2/2	3/3
Computation time	SPPC	0.0228s	0.0222s	0.0220s	0.0220s
	EPPC	0.144s	0.151s	0.148s	0.145s
Mean RMS error in mm	SPPC	0.21	1.08	2.28	3.47
	EPPC	0.20	0.97	1.97	2.94
Standard deviation of the mean RMS error	SPPC	0.08	0.43	0.97	1.51
	EPPC	0.07	0.37	0.76	1.12
Maximal RMS error in mm	SPPC	0.55	3.09	6.79	9.84
	EPPC	0.55	2.72	5.92	7.47
Ratio of wrong convergence	SPPC	21.9%	21.0%	19.8%	18.4%
	EPPC	0.26%	0.08%	0.02%	0.14%
Relative error		1.085	1.096	1.117	1.130
Ratio of success		7249	7103	7165	7149

Table 8: Influence of the noise on the robustness.

Table 9 investigates the influence of the number of points used in the registration. The RWC is still negligible for the EPPC whereas it decreases from 21.2% to 15.8% as the number of point decreases.

Finally, Table 10 presents almost the same conclusions for the influence of the angle between the cameras. Indeed, the RWC for the EPPC is always under 0.5% and the RWC for the SPPC decreases from 23.6% to 15.9% with the angle.

To summarize, the SPPC converges wrongly in broadly 20% of the case, whereas the EPPC almost always converges toward the optimal transformation. In addition, we can point out that the EPPC issues a more accurate solution than the short one in 70% of all the case. Therefore, we can conclude that the EPPC is a very robust criterion for our kind of application and that its basin of convergence is much wider than the SPPC one. However, we have to remain careful that the robustness results depend on the optimization's method as well. Therefore another optimization method might lead to slightly different results.

Number of points		30	15	8	4
Computation time	SPPC	0.0385s	0.0221s	0.0137s	0.0098s
	EPPC	0.411s	0.148s	0.066s	0.034s
Mean RMS error in mm	SPPC	1.60	2.26	2.86	4.62
	EPPC	1.42	1.96	2.52	3.93
Standard deviation of the mean RMS error	SPPC	0.66	0.95	1.17	1.94
	EPPC	0.55	0.75	0.94	1.51
Maximal RMS error in mm	SPPC	4.58	6.78	7.79	12.10
	EPPC	3.81	5.18	7.12	10.02
Ratio of wrong convergence	SPPC	21.2%	19.8%	17.6%	15.8%
	EPPC	0.21%	0.03%	0.06%	0.04%
Relative error		1.109	1.117	1.106	1.143
Number of success		7184	7110	7034	7073

Table 9: Influence of the number of points on the robustness.

Angle between the cameras		90	60	45	30	10
Computation time	SPPC	0.0217s	0.0218s	0.0220s	0.0224s	0.0222s
	EPPC	0.110s	0.131s	0.149s	0.167s	0.179s
Mean RMS error in mm	SPPC	1.84	2.05	2.27	2.52	4.41
	EPPC	1.77	1.84	1.97	2.19	3.49
Standard deviation of the mean RMS error	SPPC	0.73	0.89	0.96	1.04	2.01
	EPPC	0.68	0.70	0.75	0.84	1.40
Maximal RMS error in mm	SPPC	4.86	6.76	6.76	6.75	12.38
	EPPC	4.75	5.42	5.93	5.99	9.72
Ratio of wrong convergence	SPPC	23.6%	20.3%	19.7%	19.0%	15.9%
	EPPC	0.08%	0.04%	0.06%	0.14%	0.10%
Relative error		1.030	1.077	1.121	1.127	1.204
Number of success		6703	6891	7225	7271	7251

Table 10: Influence of the angle on the robustness.

Random Initialization with Identity T: We present briefly in this paragraph a symmetric version of the previous experiment: the sought transformation T is now the identity while the initial transformation is randomly chosen. Thus, one would expect approximately the same result for the RWC values and the computations time.

In fact, Table 11 presents a very different behavior: the computations time are smaller (0.100 sec. instead of 0.155 sec. for the EPPC) and the RWC is exactly nul for the EPPC and equal to 3.35% instead of 19.8% for the SPPC. It is very likely that the optimization method converges more easily toward a nul value (since the rotation is parameterized by a rotation vector).

This might be due to the optimization method, or to the particular representation of the rotation space. It would be interesting to perform some experiences with a different optimization method and with a different rotation parameterization to clarify this phenomenon. The next experiment, where both the sought and initial transformation are randomly chosen, seems to confirm the computation times and robustness values previously obtained, thus showing that there is a bias of our optimization method toward the identity.

Computation time	SPPC	0.0111s
	EPPC	0.097s
Mean RMS error in mm	SPPC	2.17
	EPPC	1.95
Standard deviation of the mean RMS error	SPPC	0.85
	EPPC	0.74
Maximal RMS error in mm	SPPC	6.65
	EPPC	5.72
Number of wrong convergence	SPPC	3.35%
	EPPC	0.0%
Relative error		1.096
Number of success		6492

Table 11: Robustness assessment with random initialization and identity T.

Random T with Random Initialization: Table 12 presents the robustness results when both T and the initialization are randomly chosen. If we compare these results with those obtained when the initialization is fixed to the identity, we note that the computations time of the SPPC changes very slightly, but that the RWC increase from

19.8% to 31.55%. Considering the EPPC, we observe an increase of the computations time from 0.15 sec. to 0.34 sec. and of the RWC from 0.03% to 0.67%. Despite a much larger RWC, it is still under 1.0%. Therefore, even if our initialization is thoroughly wrong, the convergence will be ensured in more than 99% of the cases.

Computation time	SPPC	0.0250s
	EPPC	0.342s
Mean RMS error in mm	SPPC	2.34
	EPPC	1.97
Standard deviation of the mean RMS error	SPPC	1.03
	EPPC	0.75
Maximal RMS error in mm	SPPC	6.75
	EPPC	5.56
Number of wrong convergence	SPPC	31.55%
	EPPC	0.67%
Relative error		1.139
Number of success		7526

Table 12: Robustness assessment with random initialization and random T.

4.1.5 Synthesis

As far as the accuracy is concerned, the EPPC always provides an RMS error better than the SPPC, but at the cost of a 10 to 20 times larger computational time. It is worth noticing that the gain in accuracy varies sharply with respect to the angle between the cameras and to the standard deviation ratio σ_{3D}/σ_{2D}. Indeed, if the angle is above 30° and if the standard deviation on the 3D data is 2 times less than on the 2D data, the EPPC brings very small improvements (relative error $\simeq 1.03$). On the contrary, with an angle of 20° and with σ_{3D}/σ_{2D} equal to 2, the relative error becomes very significant ($\simeq 1.30$). In an intermediate cases with less constraints (angle $\simeq 45°$ and $\sigma_{3D}/\sigma_{2D} \simeq 1$), we can rely on a relative error around 1.10.

Considering now the robustness aspect, the study showed that the EPPC is much more robust than the SPPC in most of the cases (RWC of less than 1% against around 20% with a random T and a fixed identity initialization). However, if a good initialization is easily available, the EPPC brings little enhancement.

Then, the choice of the criterion will rely heavily on the constraints of the application. In a larger context where temporal constraints would not be heavy, we recommend the EPPC because it provides a better RMS error. Furthermore, it permits a wider initialization without disturbing the quality of the registration accuracy.

Angle between the cameras		60^o	40^o	10^o
Computations time	SPPC	0.023 s	0.025 s	0.027 s
	EPPC	0.21 s	0.21 s	0.25 s
Mean TRE RMS error in mm	SPPC	1.80	1.99	2.30
	EPPC	1.70	1.83	2.20
Relative error		1.062	1.089	1.054

Table 13: Evaluation of the accuracy and computation times on real data. EPPC is up to 9% more accurate than SPPC for a computation time 10 times higher.

4.2 Evaluation with Real Data

To get closer to real clinical conditions, we present in this section some accuracy experiments on real 3D and 2D images of a phantom of the human abdomen. Around 40 radio-opaque markers are stick on its soft skin, and 2D and 3D localizations are made interactively in the CT and video images.

Ideally, the TRE should be assessed by comparing each registration result with a gold-standard that relates both the CT and the camera coordinate systems to the same physical space, using an external and highly accurate apparatus. As such a system is not available (if not we would not have to develop a new 3D/2D criterion), we adapted the registration loops protocol introduced in [PGT98, RPMA01, PBH+01], that enables to measure the TRE error vector for a given set of test points. A full description of this measure protocol can be found at Sec. 5.3 and the experimental setup is further detailed in [NPSA03]. Approximately 25 stick fiducials among the 40 were randomly chosen in each image and the registration algorithms were carried out with $\sigma_{3D} = 0.75$ mm and $\sigma_{2D} = 2.0$ pixel. This lead to the quantitative evaluation in Table 13.

One can observe that EPPC is always more accurate than SPPC but the relative error does not increase as much as denoted with synthetic data when the angle is small (10^o). This may be explained by the various conditions in which the measurement were done (different local lengths, number of fiducial observed...), and the simple representation of the camera (pinhole model).

Another explanation is the consistent but non-rigid motion (mentioned in Sec. 5.3.2) of the skin markers on which the registration is done. However, these small movements realistically simulate the imperfect repositioning of the skin and the organs that is induced by our gas volume monitoring protocol. In this context, the assumption (on which are based our criteria) of independent noises on each 3D

marker position may not be fulfilled. Despite these variations from the theoretical assumptions, we underline that the achieved accuracy is around 2mm, which is by far better than the 5mm needed for our medical application. One can assess the visual accuracy of our registration on one particular case in Fig. 1.

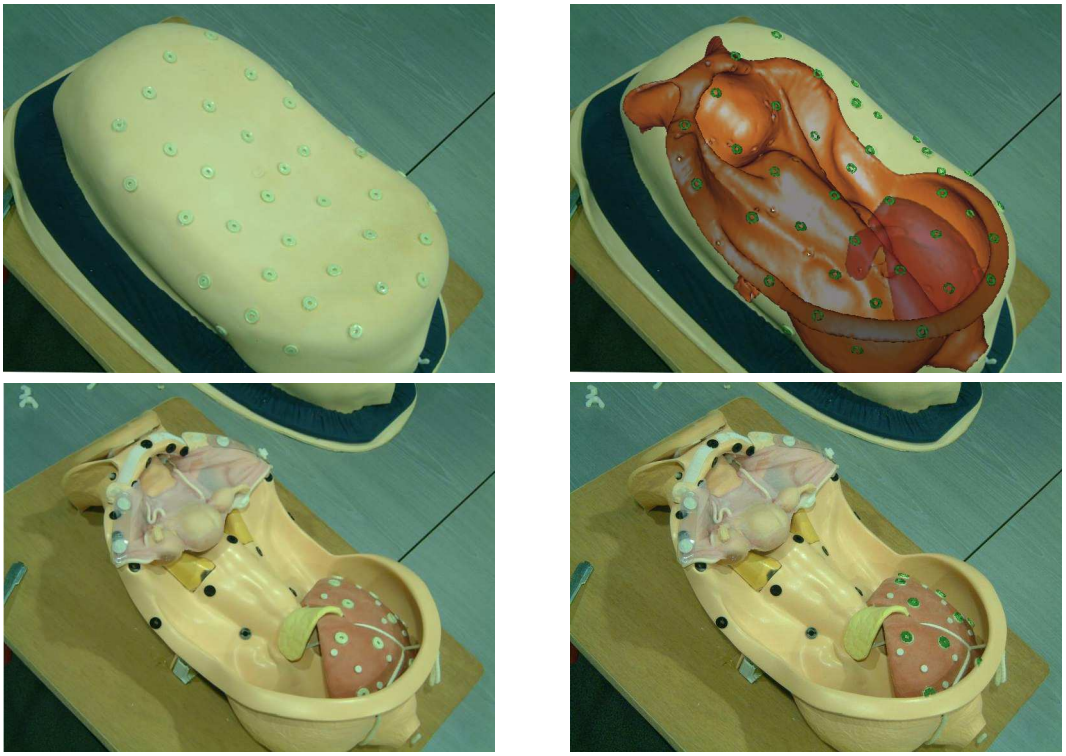

Figure 1: The top left image shows the plastic mannequin with the radio-opaque fiducials. The top right image displays the augmented reality view of the surgeon, i.e. the super-imposition on the video image of the 3D reconstruction of the fiducials and internal parts (plastic frame and liver) after the registration process. To check visually the quality of the achieved registration on the liver, we put off the skin (bottom left image), and we superimposed the reconstruction of the fiducials on the liver (bottom right).

5 Validation of the Uncertainty Predictions

In the tables 1,2,3,5,6 of the section 4.1.3, one can clearly see that the mean predicted RMS error (referred from now on to the PTRE for *Predicted* TRE) closely fits the measured mean RMS error (called now ETRE for *Experimental* TRE) for all the investigated configurations of points, noise and cameras angle. Then, it seems that our error prediction is very close to the reality. However, in these experiments, we only ensured that our prediction was accurate on a few particular cases *with synthetic conditions*. In fact, there might be some inconsistencies between the real data and the statistical assumptions used to derive the theoretical formula (small non-linearity of the criterion, perfect calibration, unbiased Gaussian noise on points, etc.) that could produce systematic biases. The goal of this section is to check incrementally that these assumptions hold within our application domain. This will be done using synthetic data (for the non-linearities of the criterion), real video images of a precisely defined 3D object (for camera calibration and distortions), and finally real CT and video images of a soft phantom of the abdomen (for noise assumptions on point measurements).

5.1 Validation on Synthetic Data

5.1.1 Methodology

To show that the non-linearities of the criterion do not induce a significant bias in our error prediction, we conduct experiments on synthetic data built as before (Sec. 4.1.1). But this time, we randomly choose the following features so that they span their whole range of variation:

- The number of point is chosen uniformly between 7 and 25.
- The position of each point is chosen between 40 different available positions which are representative of a human chest.
- The angle between the camera is an uniform variable with a minimum value of 5° and a maximum of 120°.
- The ratio of the focal distance on the distance between the cameras and the object varies from 20 to 50.
- σ_{2D} and σ_{3D} are chosen uniformly between 0.5 and 4.0 (which corresponds to a SNR of 60 dB to 90 dB[3]).

[3] $SNR_{dB} = 10 \log_{10}(\frac{\sigma_s}{\sigma_n})$ where σ_s (resp. σ_n) is the variance of the signal (resp. noise).

Moreover, as our error prediction only holds when the algorithm has correctly converged and since we are not interested here in robustness evaluation, we initialize the algorithm with a transformation sufficiently close to the right solution.

Since each experiment is different, we need to evaluate the relative fit of the PTRE vs. the ETRE to quantitatively measure the quality of the uncertainty prediction. Due to the significant anisotropy, we did not use the basic ratio $ETRE^2/PTRE^2$, but rather the statistical framework introduced in [PT97] and its validation index, which weights the observed error vector with the inverse of its predicted covariance matrix to yield a Mahalanobis distance μ^2. More specifically, if the real error after registration on the test points follows a Gaussian distribution with null mean and the predicted covariance, then the validation index follows a χ^2_3 law. Repeating this experiment with many different "parameters" configurations to obtain many *independent* values of μ^2, we can verify that it is really χ^2_3 distributed. The Kolmogorov-Smirnov test [PFTV91] is well adapted to do that (referred from now on to the K-S test), but since it only gives a binary answer, we also use the fact that the empirical mean value of a χ^2_3 distribution is 3 and its variance is 6. We call *validation index* the estimated mean value of μ_i^2: $I = \bar{\mu}^2 = \frac{1}{N}\sum_{i=0}^{N} \mu_i^2$ and its variance is computed with $\sigma_I^2 = \frac{1}{N-1}\sum_{i=0}^{N}(\mu_i^2 - \bar{\mu}^2)^2$.

This index can be interpreted as an indications on how the estimation method under-estimates ($I > 3$) or over-estimates ($I < 3$) the error on the test points. It is a kind of relative error on the error estimation.

To summarize, we compute for each registration the predicted error on the transformation and we propagate it on the 9 test points C_i. However, the test points being located in the same region, their covariances matrices are highly correlated. Thus, to fulfill the independence condition of the K-S test, we need to choose one test point (randomly among the 9) per experiment.

5.1.2 Validation

Figures 2 to 6 show the validation index and the KS-confidence for 5 independently varying parameters. Each value of the validation index and corresponding K-S test is computed on the result of 1000 registrations. These figures clearly show that the prediction of the transformation uncertainty is always truthful. Indeed, the validation index is always very close to 3.0 and the KS-confidence values never fall under 1%. This means that there is little chance that the resulting transformation

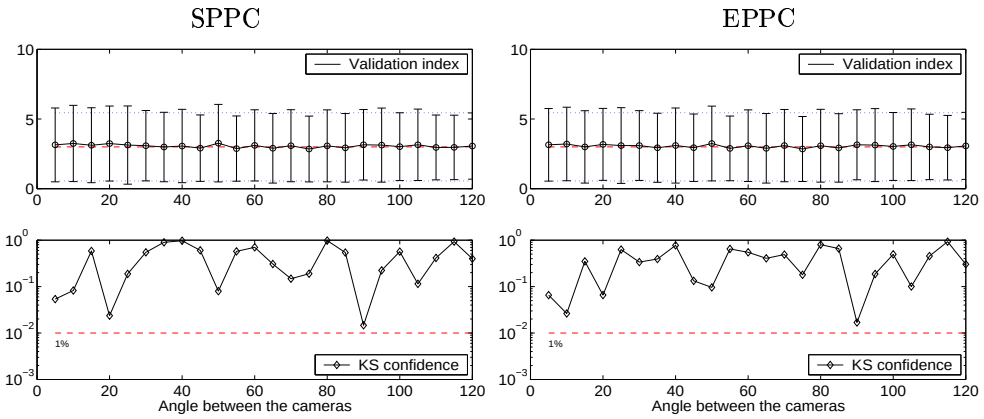

Figure 2: Validation of the uncertainties prediction w.r.t. the angle between the cameras. Top image: mean and standard deviation of the synthetic validation index. Bottom image: KS confidence. It validates in the whole case with a confidence above 1%.

noise does not behave like the Gaussian distribution we predict. We can just point out that for a number of points smaller than 7, the index validation is around 3.2, which means we are slightly under-estimating the resulting noise on the transformation.

Table 14 summarize the statistics obtained for 20000 registrations where all five parameters randomly vary together. The figures obtained for both the validation index and the KS-test fully validate the reliability of the transformation's prediction. In conclusion, we can assess that we are able to predict accurately the transformation uncertainty (in the range of our application constraints), if the calibration process is perfect and if the covariance on the points is perfectly known. Of course, these are not realistic assumption: we still have to confront our algorithm to some real cases.

	Val. index μ^2 (3.0)	Std. dev. (2.449)	KS-conf. (0.01< x <1.0)
SPPC	3.020	2.506	0.353
EPPC	3.016	2.486	0.647

Table 14: Validation of the uncertainties prediction in the broad case after 20000 registrations.

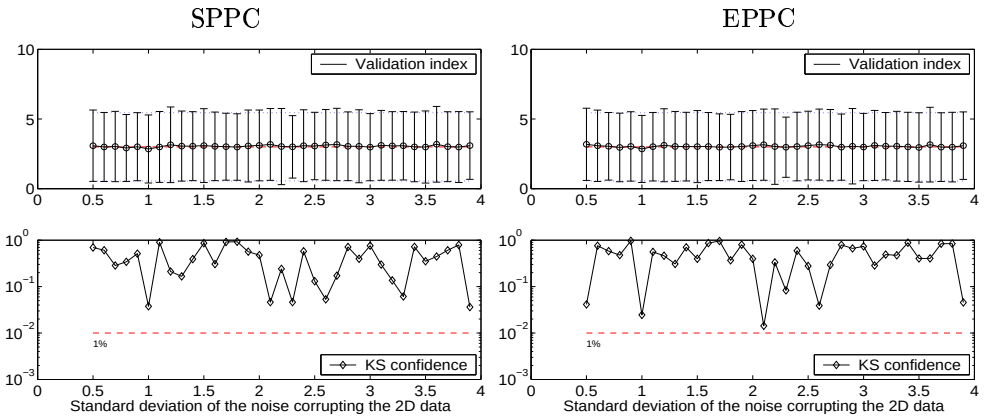

Figure 3: Validation of the uncertainties prediction w.r.t. the noise corrupting the 2D data. Top image: mean and standard deviation of the synthetic validation index. Bottom image: KS confidence. It validates in the whole case with a confidence above 1%.

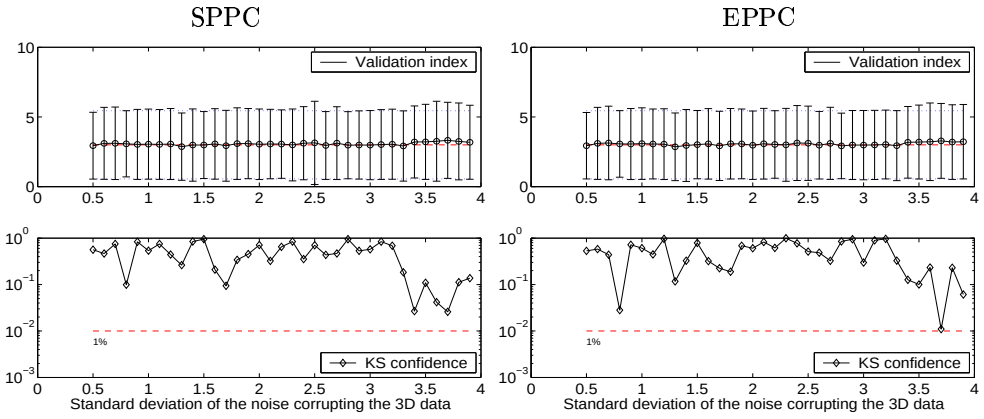

Figure 4: Validation of the uncertainties prediction w.r.t. the noise corrupting the 3D data. Top image: mean and standard deviation of the synthetic validation index. Bottom image: KS confidence. It validates in the whole case with a confidence above 1%.

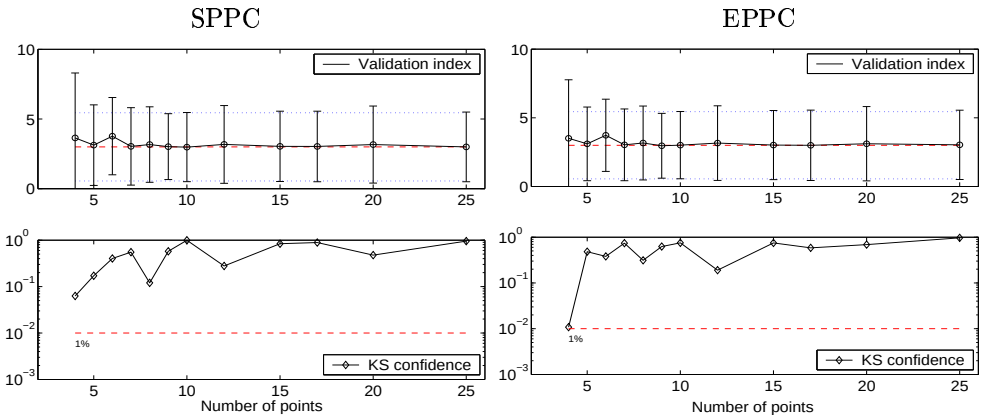

Figure 5: Validation of the uncertainties prediction w.r.t. the number of points used for the registration. Top image: mean and standard deviation of the synthetic validation index. Bottom image: KS confidence. It validates in the whole case with a confidence above 1%.

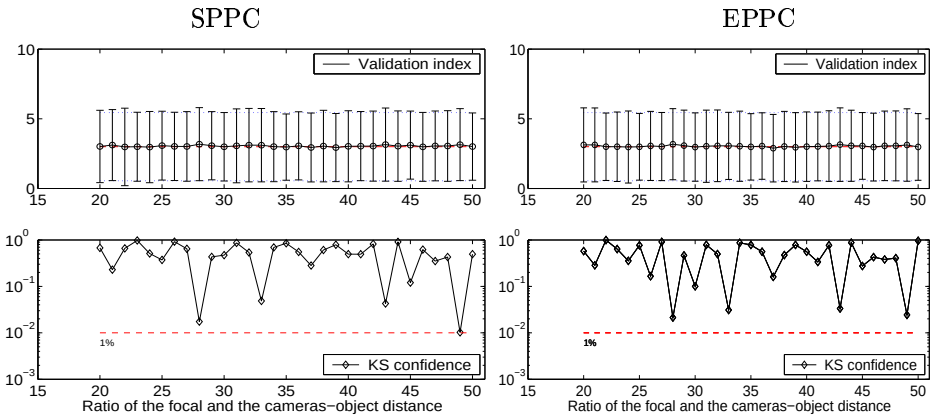

Figure 6: Validation of the uncertainties prediction w.r.t. the ratio focal/distance. Top image: mean and standard deviation of the synthetic validation index. Bottom image: KS confidence. It validates in the whole case with a confidence above 1%.

5.2 Validation with real Calibration and Synthetic Noise

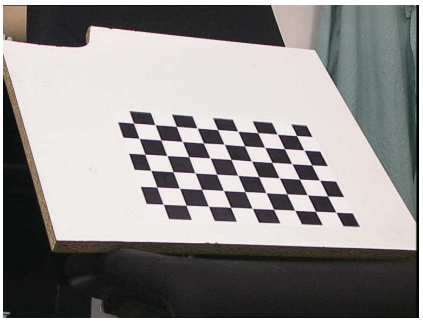

Figure 7: Calibration grid used for 3D/2D registration.

The perfect validation of our accuracy prediction on synthetic data does not take into account possible calibration errors of the cameras and excludes likely distortions from the pinhole model. The goal of this experiment is to address the validity of these assumptions using a real video system. We used a 54 points calibration grid (see Fig. 7) that allows for a very accurate detection of the points ($\sigma_{3D} < 0.1$ mm, $\sigma_{2D} < 0.2$ pixel). Such an accuracy is obviously far below the current detection of real markers positions ($\sigma_{2D} \simeq 2$ pixel, $\sigma_{3D} \simeq 1$ mm). Adding an artificial but realistic noise to the collected data, we validate the error prediction *when calibration is real*, but still under the assumptions of a centered Gaussian noise.

5.2.1 Methodology

As mentioned in Sec. 4.2 we should assess the ETRE by comparing each registration result with a gold-standard that is not available. To circumvent the lack of gold-standard, we adapted the registration loops protocol introduced in [PGT98, RPMA01, PBH+01], that enables to measure the TRE error vector for a given set of test points and enables validation of the error prediction with the validation index previously introduced in the section 5.1.

The principle is to acquire several couples of 2D images with jointly calibrated cameras so that we can compare 2 *independent* 3D/2D registrations $\hat{T}_1$ and $\hat{T}_2$ of the same object (each registration is computed with different 2D and 3D images). A typical loop, sketched in Fig. 10, described the method to get a μ^2-value. We should observe that $E = \hat{T}_1^{-1} \cdot \hat{T}_2$ is very close to the identity matrix. Computing the covariance matrix of each transformation of the loop, we can propagate the noise from the input data and compute the 3×3 covariance matrix $\Sigma_{\hat{T}_1^{-1}\cdot\hat{T}_2\cdot C}$ of the target point C after the loop. Consequently, we can compute the following validation index:

$$\bar{\mu}^2 = 1/N \cdot \sum_{i=1}^{N} \mu_i^2 = 1/N \cdot \sum_{i=1}^{N} (C - E \star C)^\top \cdot \Sigma^{-1}_{\hat{T}_1^{-1}\cdot\hat{T}_2\cdot C} \cdot (C - E \star C)$$

and check that μ^2 follows a χ^2_3 law[4].

This experiment providing only one error measurement, we still need to repeat it with different datasets to obtain statistically significant measures. In order to take into account possible calibration error and/or bias, it is necessary to change the cameras calibrations and positions, and not only to move the object in the physical space. Likewise, to decorelate the two 3D/2D transformations, we need to use two differently noised 3D data sets. Indeed, when using the same set of 3D points to register the 2D points, the error on 3D points similarly affects both transformations, and the variability of the 3D points extraction (and any possible bias) is hidden.

Finally, for each set of parameters (different configuration of our four cameras, different positions/orientations of the calibration grid), we got 144 μ^2-values. The cameras were placed 10^o to 60^o apart, at a distance of the object of 25 to 30 times the focal length.

5.2.2 Validation

Figures 8 shows the mean, standard deviation and K-S test value of the validation index w.r.t. the number of points used (randomly chosen among the 54 available). One can see that the prediction is correct up to 40 points (which spans our range of application). This critical value is due to the progressive reduction of the registration error that finally meets the ignored calibration error (about 0.5 mm). Likewise, we observed on Fig. 9 the same behavior when the feature noise becomes too small (σ_{3D} and σ_{2D} below 0.7).

5.3 Validation with Real Data

5.3.1 Methodology

To test the last assumption of our prediction (unbiased Gaussian noise), we now turn to a validation experiment on real 3D and 2D measurements of the plastic phantom, on which are stick about 40 radio-opaque markers. The set up is slightly different from the one used for the previous calibration grid. Firstly, target points C_i are no more chosen at a certain distance of the grid but within the phantom liver. Secondly, we used several different CT-images to decorelate the two 3D/2D transformations. Indeed, if we register the 2D points to the same set of 3D points extracted from a

[4]In fact, we can perform the same validation process directly with the transformation uncertainty prediction. Indeed, we can compute the uncertainty of the two registrations composition $\Sigma_{T\ loop}$. In such case, the validation index has this form: $\bar{\mu}^2 = 1/N \cdot \sum_{i=1}^{N} E^\top \cdot \Sigma^{-1}_{T\ loop} \cdot E$ and μ^2 follows a χ^2_6 law if our prediction is reliable.

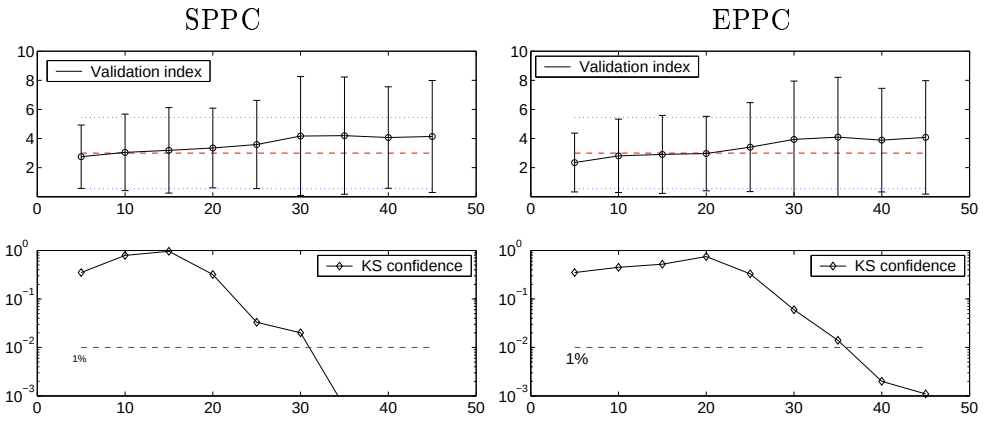

Figure 8: Validation of the uncertainty prediction on the calibration grid w.r.t. the number of points used for the registration. Noise is set to $\sigma_{2D} = 2$ pixels and $\sigma_{3D} = 2$mm. Top: mean and standard deviation of the validation index. Bottom: KS confidence. It validates up to 40 points with a confidence above 1%

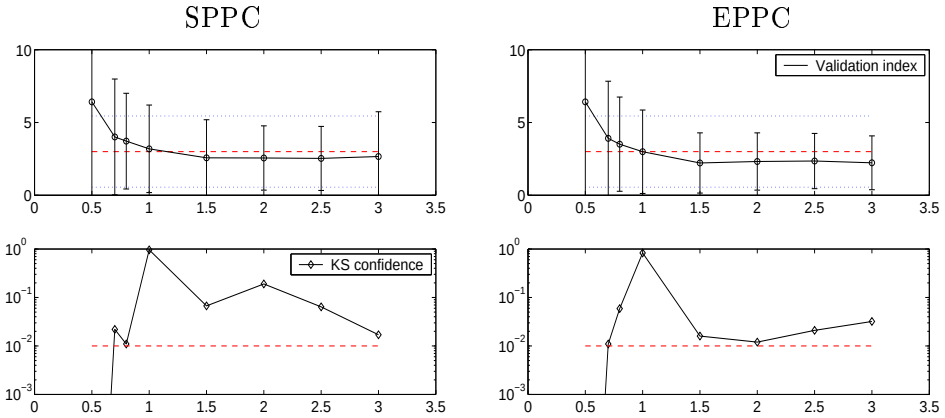

Figure 9: Validation of the uncertainty prediction on the calibration grid w.r.t. noise corrupting the 2D and 3D data. Number of points used for registration is 15. Top: mean and standard deviation of the validation index. Bottom: KS confidence. It validates with a confidence above 1% till σ_{2D} and σ_{3D} are above 0.7

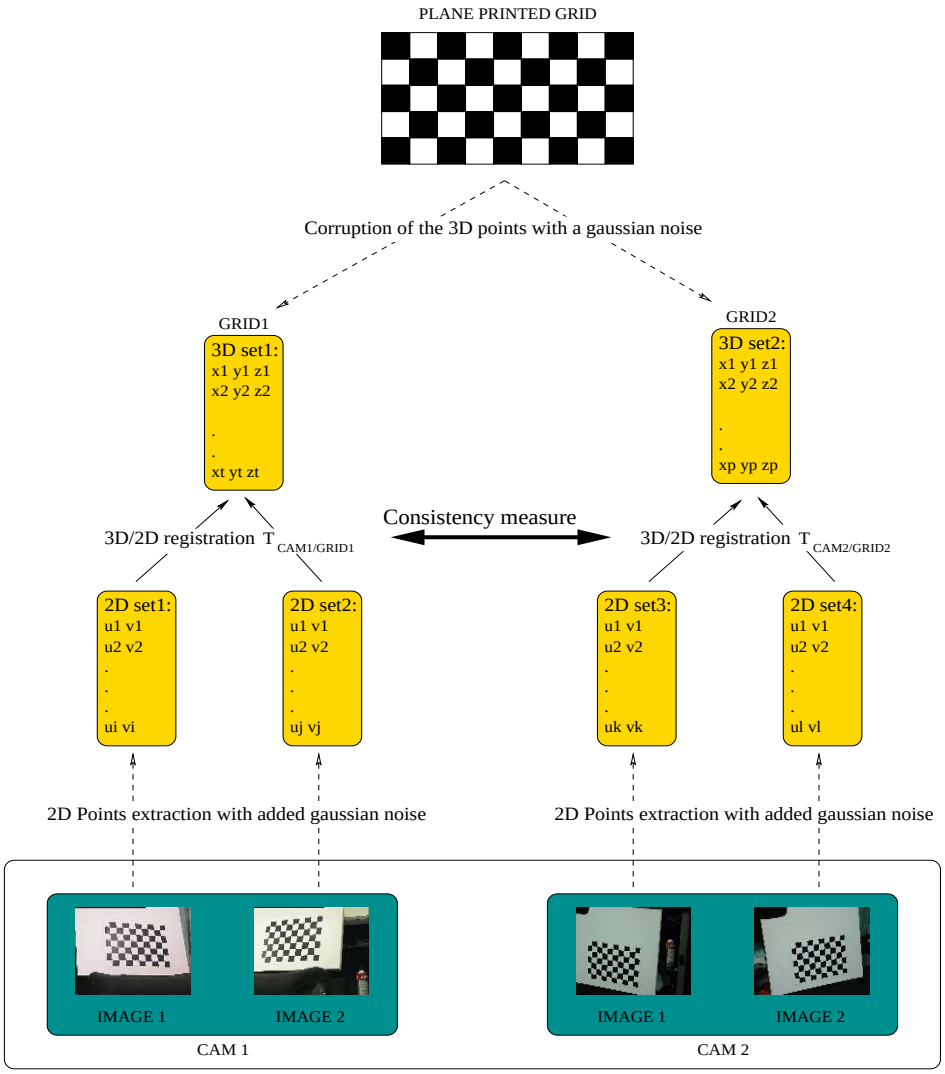

Figure 10: Registration loops used to estimate the registration consistency: a test point C chosen at a certain distance of the printed grid (typically 20 cm) is transformed into the CAM1 coordinate system using a first 3D/2D registration $\hat{T}_1$, then back into the grid coordinate system using a second 3D/2D registration $\hat{T}_2$ provided by the other couple of cameras (the coordinate system of CAM1 and CAM2 are identical since cameras are jointly calibrated). If all transformations were exact, we would obtain the same position for the test point. Of course, since the transformations are not perfect, we measure a non-isotropic 3D variance vector ${\sigma^2}_{loop} = 2{\sigma^2}_{CAM/GRID}$ that corresponds to a TRE. Finally, a μ^2-value is given by the Mahalanobis distance between C and $\hat{T}_1 \star \hat{T}_2^{-1} \star C$.

single CT scan, the two transformations will be similarly affected by the 3D points. Consequently, the variability of the 3D points extraction (and any possible bias on it) will be hidden and the effect of the sampling onto the markers localization will be ignored.

Incorporating 2 different CT-scans of the phantom modifies the loop (see Fig. 11) so that we now have $\sigma^2_{loop} = 2{\sigma^2}_{CAM/CT} + {\sigma^2}_{CT/CT}$. Consequently, if we want to assess faithfully $\sigma_{CAM/CT}$, we need to have an accurate estimation of $\sigma_{CT/CT}$, or to carry out almost perfect intra-modal registration (in which case the intra-modal term will be negligible in the formula). To solve this problem we devise an experimental procedure described in the Sec. 5.3.2 based on multiple CT image registration that lead to a bronze standard registration between the CT images.

The markers used were randomly chosen among the 40 available, and the cameras were placed 20^o to 60^o apart, at a distance of the object of 25 to 30 times the focal length. For each experiment, we obtained 80 μ^2-values. As we experimentally observed that there was a consistent but non-rigid motion of the soft skin (about 1mm), we chose $\sigma_{3D} \simeq 2.0$ mm (instead of 1 mm) to take into account this additional uncertainty.

5.3.2 Bronze standard Registration between the CT images

Our goal here is to compute the $n-1$ most reliable transformations $\bar{T}_{i,i+1}$ that relate the n (successive) CT_i images. Estimations of these transformations are readily available by computing all the possible registrations $T_{i,j}$ between the CT images using m different methods ([RPMA01]). Then, the transformations $\bar{T}_{i,i+1}$ that best explain these measurements are computed by minimizing the sum of the squared distance between the observed transformations $T_{i,j}$ and the corresponding combination of the sought transformation $\bar{T}_{i,i+1} \circ \bar{T}_{i+1,i+2} \ldots \bar{T}_{j-1,j}$. The distance between transformations is chosen as a robust variant of the left invariant distance on rigid transformation developed in [PGT98].

The estimation $\bar{T}_{i,i+1}$ of the perfect registration $T_{i,i+1}$ is called *bronze standard* because the result converges toward $T_{i,i+1}$ as m and n become larger. Indeed, considering a given registration method, the variability due to the noise in the data decreases as the number of images n increases, and the registration computed converges toward the perfect registration up to the intrinsic bias (if there is any) introduced by the method. Now, using different registration procedures based on different methods, the intrinsic bias of each method also becomes a random variable, which is hopefully centered around zero and averaged out in the minimization procedure. The different bias of the methods are now integrated into the transformation variability. To fully

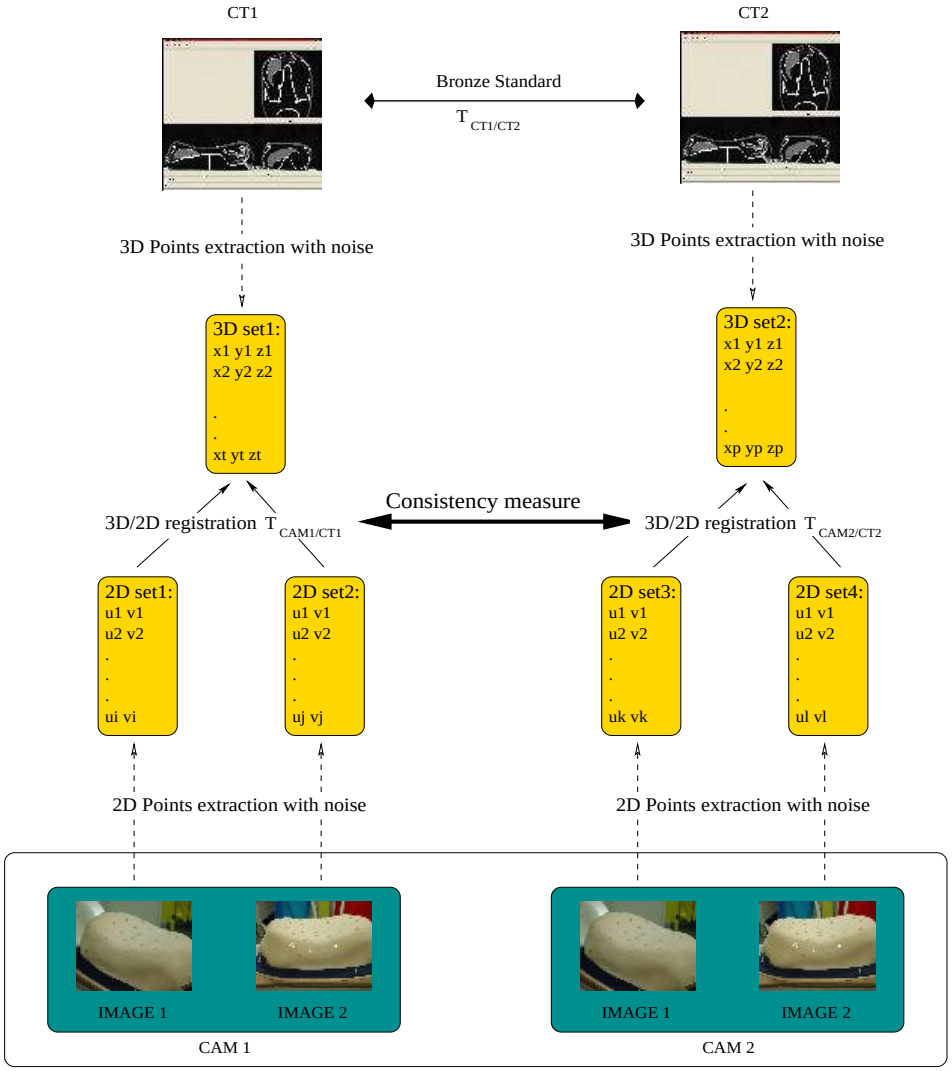

Figure 11: Registration loops used to estimate the registration consistency: a test point C chosen inside the phantom's liver and defined in the CT1 coordinate system is transformed into the CAM1 coordinate system using a first 3D/2D registration $\hat{T}_1$, then into the CT2 coordinate system using a second 3D/2D registration $\hat{T}_2$ provided by the other couple of cameras (the coordinate system of CAM1 and CAM2 are identical since cameras are jointly calibrated). Finally the test point is transformed back into the CT1 coordinate system thanks to our bronze standard registration. If all transformations were exact, we would obtain the same position for the test point. Of course, since the transformations are not perfect, we measure a non-isotropic 3D variance vector ${\sigma^2}_{loop} = 2{\sigma^2}_{CAM/CT} + {\sigma^2}_{CT/CT}$ that corresponds to a TRE. Finally, a ${\mu^2}$-value is given by the Mahalanobis distance between C and $\hat{T}_1 \star \hat{T}_2^{-1} \star \hat{T}_{CT2/CT1} \star C$.

reach this goal, it is important to use as many independent registration methods as possible.

In our setup, we used five CT scan of the plastic mannequin in different positions, and five different methods with different geometric features or intensity measures. Three of these methods are intensity-based: the algorithm `Aladin` [ORPA00] has a block matching strategy where matches are determined using the coefficient of correlation, and the transformation is robustly estimated using a least-trimmed-squares; the algorithm `yasmina` uses the Powell algorithm to optimize the SSD or a robust variant of the correlation ratio (CR) metrics between the images [RPMA01]. For the feature-based methods, we used the `crest lines` registration described in [PAT00], and the multi-scale `EM-ICP` algorithm of [GP02] on zero-crossings of the Laplacian surfaces (the images were down-sampled by a factor of 2 to limit the number of surface points to about 1.5 million...). Since none of these methods uses the 3D extracted points as registration data, we ensure the independence with respect to the marker localization noise that corrupts the 3D/2D registration.

As a side effect, we may use only four of the five methods to determine the bronze standard registration, and use that standard to determine the accuracy of the fifth method (a kind of leave-one-method-out test). This uncertainty is then propagated into the final bronze standard registration (including all methods) to estimate its accuracy. In the table below, we give the standard deviation determined this way on the rotational and translational components of each method, the uncertainty of the resulting bronze standard registration and the uncertainty of the 3D registration using standard least-squares of the fiducial markers positions (w.r.t. the bronze standard).

	σ_{rot} (deg)	σ_{trans} (mm)
Aladin	0.09	0.56
Yasmina SSD	0.02	0.41
Yasmina CR	0.06	0.41
Crest lines	0.04	0.27
EM-ICP	0.08	0.68
Bronze standard	0.01	0.07
Fiducials	0.15	0.85

One can observe that the crest lines registration is performing the best, quickly followed by the Yasmina registrations. EM-ICP is not performing very well due to the down sampling of the images. The final bronze standard registration accuracy is very good (it corresponds to 0.08 mm TRE on the test points). Finally, we point

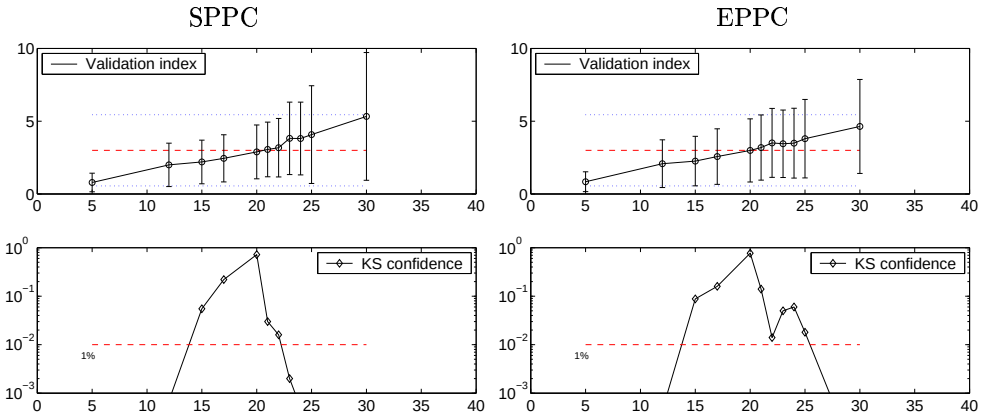

Figure 12: Validation of the uncertainty prediction on the phantom w.r.t. the number of points used for the registration. Top: mean and standard deviation of the validation index. Bottom: KS confidence.

out that the lack of accuracy of the fiducials registration w.r.t. the other methods is due to the fact that the markers were stick on the "skin" of the mannequin, which is elastic and did move by about 1 to 2 mm between the acquisitions, while all other methods did focus on the rigid structure of the mannequin (this effect was checked on the images after registration).

5.3.3 Validation

Figure 12 presents the mean and variance of μ^2 w.r.t. the number of points. Firstly, we notice that the mean value slowly increases with the number of points. This can be explained by the biases introduced by the calibration error and the correlated motion of the markers on the skin. Indeed, the measured accuracy figures do not converge to 0 mm with a large number of points but rather towards 1 mm, which corresponds to the motion of the skin.

Nevertheless, it appears that the prediction is well validated for a range of 15 to 25 points. As $\bar{\mu}^2$ can be interpreted as a relative error or the error prediction (see [PT97]), Fig. 12 shows that we over-estimate the mean TRE by a factor 1.7 for a small number of points ($\bar{\mu}^2 \simeq 1$), and that we under-estimate it by a factor of 1.3 for more than 25 points ($\bar{\mu}^2 \simeq 5$). For our application, in which the number of visible points should not exceed 20, this means that we predict correctly the amplitude of the error on the transformation. In the worst case, we over-estimate it, which can be considered as a good safety measure.

6 Conclusion

We devised in this paper an augmented reality system for RF guidance based on a new 3D/2D registration criterion with a validated error prediction. To reach the best accuracy performances, we firstly derived a new 2D/3D Maximum Likelihood registration criterion (EPPC) based on better adapted statistical hypotheses than the classical 3D/2D least-square registration criterion (SPPC). Experiments on synthetic and real data showed that EPPC is 5 to 20% more accurate and much more robust than SPPC, but requires higher computation times when the initialization is unknown (0.3 instead of 0.03 sec.). However, synthetic experiments provide evidences that a refreshment rate of 20 to 40 Hz is achievable in a tracking phase, where only small motions have to be detected. In the context of augmented reality for RF liver tumors ablation, we showed that the proposed method provides an accuracy of about 2mm within the liver, which fits the initial goal of 5mm that was necessary to provide a significant help for radiologists and surgeons. We underline an alternative interpretation of this gain: we can typically reach the same accuracy with 20 markers for EPPC where 24 are needed for SPPC. As we face possibilities of markers occlusion because of the surgeon's hand and cumbersomeness constraints on the placement of the markers, this gain has to be considered seriously. In addition, as clinical conditions do not allow a free camera positioning, we could meet situation where an angle between the cameras could decrease below 20^o, which would mean an accuracy gain of 18%.

In order to assess the system accuracy for all configurations, we propose in a second step a theoretical propagation of the target covariance through SPPC and EPPC w.r.t the experimental configuration parameters. To verify the validity of all the assumptions of that method, we conducted a careful validation study that assess in turn the range of validity of each hypothesis. We firstly verified that non-linearities in the criterion and calibration error are negligible. Then, we use a realistic phantom with a soft and deformable skin to validate the prediction in the range of our application (i.e. for 15 and 25 markers). This study confirmed that we correctly predict the registration error, with a slight over-estimation if too much markers are occluded, which is a good safety rule.

To reach the clinical usability, the whole system still has to be validated on real patients. We are currently conducting experiments (using repeated CT scans at the same point of the breathing cycle) to certify that the motion of the internal structures due to the monitored breathing of the patient cannot bias our accuracy prediction. Preliminary results indicates that this motion is of the order of 1 mm, which is in accordance with the motions we experienced because of the phantom soft

skin. Thus, we are pretty confident that our registration error prediction will work properly in the final system. Last but not least, it is possible to estimate broadly the TRE *before* scanning the patient, by using the stereoscopic reconstruction of the markers instead of their positions in the scanner. This will allow a better control of the external conditions (number of markers, angle between the cameras) and the optimization of the intervention preparation.

Acknowledgments This work has been supported by INRIA and IRCAD through the PhD fellowship of S. Nicolau.

References

[BLM+98] J.M. Balter, K.L. Lam, C.J. McGinn, T.S. Lawrence, and R.K. Ten Haken. Improvement of ct-based treatment-planning models of abdominals targets using static exhale imaging. *Int. J. Radiation Oncology Biol. Phys.*, 41(4):939–943, 1998.

[DD95] D. DeMenthon and L. Davis. Model-based object pose in 25 lines of code. *International Journal of Computer Vision*, 15:123–141, 1995.

[DRL89] M. Dhome, M. Richetin, and G. Lapreste, J.and Rives. Determination of the attitude of 3d objects from a single perspective view. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(12):1265–1278, December 1989.

[FB81] M. Fischler and R. Bolles. Random sample consensus : A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, June 1981.

[Gan84] S. Ganapathy. Decomposition of transformation matrices for robot vision. *Pattern Recognition Letters*, 2(6):401–412, December 1984.

[GEW+96] W. Grimson, G. Ettinger, S. White, W. Wells T. Lozano-Perez, and R. Kikinis. An automatic registration method for frameless stereotaxy, image-guided surgery and enhanced reality visualization. *IEEE Trans. Medical Imaging*, 15(2):129–140, April 1996.

[GP02] Sébastien Granger and Xavier Pennec. Multi-scale EM-ICP: A fast and robust approach for surface registration. In A. Heyden, G. Sparr,

M. Nielsen, and P. Johansen, editors, *European Conference on Computer Vision (ECCV 2002)*, volume 2353 of *LNCS*, pages 418–432, Copenhagen, Denmark, 2002. Springer.

[Her98] Gerardo Hermosillo. Recalage par maximisation de l'information mutuelle. Technical report, Université de Nice, Septembre 1998. Rapport de stage de DEA.

[HJL+89] R. Haralick, H. Joo, CN. Lee, X. Zhuang, V. Vaidya, and MB. Kim. Pose estimation from corresponding point data. *IEEE Trans. on Systems, Man. and Cybernetics*, 19(06):1426–1446, December 1989.

[HYH85] Y. Hung, PS. Yeh, and D. Harwood. Passive ranging to known planar point sets. *IEEE Int. Conf. on Robotics and Automation*, pages 80–85, March 1985.

[Kum89] R. Kumar. Determination of camera location and orientation. In *DARPA Image Understanding Workshop*, pages 870–881, Palo Alto, Calif., 1989.

[Lev97] M. Leventon. A registration, tracking, and visualization system for image guided surgery. Master's thesis, MIT, May 1997.

[LHF90] Y. Liu, T. Huang, and O. Faugeras. Determination of camera location from 2d to 3d line and point correspondences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(01):28–37, January 1990.

[Low91] D. Lowe. Fitting parameterized three-dimensional models to images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(5):441–450, May 1991.

[LWG97] M. Leventon, W. Wells, and W. Grimson. Multiple view 2d-3d mutual information registration. In *Image Understanding Workshop*, pages 625–630, 1997.

[MFW+97] C. Maurer, J. Fitzpatrick, M. Wang, R. Galloway, R. Macinuas, and G. Allen. Registration of head volume images using implantable fiducial markers. *IEEE Trans. Med. Imaging*, 16:447–462, 1997.

[NPSA03] S. Nicolau, X. Pennec, L. Soler, and N. Ayache. Evaluation of a new 3D/2D registration criterion for liver radio-frequencies guided by

augmneted reality. In N. Ayache and H. Delingette, editors, *International Symposium on Surgery Simulation and Soft Tissue Modeling (IS4TM'03)*, volume 2673 of *Lecture Notes in Computer Science*, pages 270–283, Juan-les-Pins, France, 2003. INRIA Sophia Antipolis, Springer-Verlag.

[OLWK98] Siu-Hang Or, W. S. Luk, K. H. Wong, and Irwin King. An efficient iterative pose estimation algorithm. In *ACCV*, pages 559–566, 1998.

[ORPA00] S. Ourselin, A. Roche, S. Prima, and N. Ayache. Block Matching: A General Framework to Improve Robustness of Rigid Registration of Medical Images. In *Third International Conference on Medical Image Computing And Computer-Assisted Intervention (MICCAI'00)*, pages 557–566, Pittsburgh, Pennsylvania USA, October 11-14 2000.

[PAT00] X. Pennec, N. Ayache, and J.-P. Thirion. Landmark-based registration using features identified through differential geometry. In I. Bankman, editor, *Handbook of Medical Imaging*, chapter 31, pages 499–513. Academic Press, September 2000.

[PBH+01] G. Penney, J. Blackall, D. Hayashi, T. Sabharwal, A. Adam, and D. Hawkes. Overview of an ultrasound to ct or mr registration system for use in thermal ablation of liver metastases. In *Medical Image Understandings and Analysis*, July 2001.

[Pen96] X. Pennec. *L'incertitude dans les problèmes de reconnaissance et de recalage. Application en imagerie médicale et biologie moléculaire.* PhD thesis, École Polytechnique, 1996.

[PFTV91] W.H. Press, B.P. Flannery, S.A Teukolsky, and W.T. Vetterling. *Numerical Recipices in C.* Cambridge Univ. Press, 1991.

[PFTV92] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. *Numerical Recipes in C.* Cambrige University Press, 1992.

[PGT98] X. Pennec, C.R.G. Guttmann, and J.-P. Thirion. Feature-based registration of medical images: Estimation and validation of the pose accuracy. In *Proc. of First Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI'98)*, volume 1496 of *LNCS*, pages 1107–1114, Cambridge, USA, October 1998. Springer Verlag.

[PHT95] T. Phong, R. Horaud, and P. Tao. Object pose from 2-d to 3-d points and line correspondences. *International Journal of Computer Vision*, 15:225–243, 1995.

[Pow64] M. J. D. Powell. An efficient method for finding the minimum of a function of several variables without calculating derivatives. *The Computer Journal*, 7(2):155–162, 1964.

[PT97] X. Pennec and J.-P. Thirion. A framework for uncertainty and validation of 3D registration methods based on points and frames. *Int. Journal of Computer Vision*, 25(3):203–229, 1997.

[QL99] L. Quan and Z. Lan. Linear n-point camera pose determination. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(07):774–780, July 1999.

[RPMA01] A. Roche, X. Pennec, G. Malandain, and N. Ayache. Rigid registration of 3D ultrasound with MR images: a new approach combining intensity and gradient information. *IEEE Transactions on Medical Imaging*, 20(10):1038–1049, October 2001.

[Sa01] L. Soler and al. Fully automatic anatomical, pathological, and functional segmentation from ct scans for hepatic surgery. *Computer Aided Surgery*, 6(3), August 2001.

[SB97] G. Simon and M.O. Berger. A two stage robust statistical method for temporal registration from features of various type. Technical Report 3235, INRIA Lorraine, August 1997.

[Vio95] Paul A. Viola. Alignment by maximization of mutual information. Technical Report 1548, MIT, June 1995.

[VW97] P. Viola and W.M. Wells. Alignment by maximization of mutual information. *International Journal of Computer Vision*, 24(2):137–154, 1997.

[Yua89] J. Yuan. A general photogrammetric method for determining object position and orientation. *IEEE TRansactions on Robotics and Automation*, 5(2):129–142, April 1989.

Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Unité de recherche INRIA Futurs : Parc Club Orsay Université - ZAC des Vignes
4, rue Jacques Monod - 91893 ORSAY Cedex (France)
Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)
Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399